D2.2
Lists of ethical, legal, societal and economic issues of big data technologies

e-SIDES

Ethical and Societal Implications of Data Sciences

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731873

Grant Agreement number: 731873

e-SIDES – Ethical and Societal Implications of Data Sciences

Data-driven innovation is deeply transforming society and the economy. Although there are potentially enormous economic and social benefits this innovation also brings new challenges for individual and collective privacy, security, as well as democracy and participation. The main objective of the CSA e-SIDES is to complement the research on privacy-preserving big data technologies, by analyzing, mapping and clearly identifying the main societal and ethical challenges emerging from the adoption of big data technologies, conforming to the principles of responsible research and innovation; setting up and organizing a sustainable dialogue between industry, research and social actors, as well as networking with the main Research and Innovation Actions and Large Scale Pilots and other framework program projects interested in these issues. It will investigate stakeholders' concerns, and collect their input, framing these results in a clear conceptual framework showing the potential trade-offs between conflicting needs and providing a basis to validate privacy-preserving technologies. It will prepare and widely disseminate community shared conclusions and recommendations highlighting the best way to ultimately build confidence of citizens and businesses towards big data and the data economy.

D2.2 Lists of ethical, legal, societal and economic issues of big data technologies

Work package	WP 2 – Establishment of common ground
Lead author	Bart Custers (Leiden University)
Contributing authors	Bart Custers (Leiden University) Karolina La Fors (Leiden University) Magdalena Jozwiak (Leiden University) Daniel Bachlechner (Fraunhofer ISI) Michael Friedewald (Fraunhofer ISI) Stefania Aguzzi (IDC Italy)
Internal review	Duncan Brown
Due Date	M8 (August 2017)
Date	31 August 2017
Version	1.0 (final)
Type	Report
Dissemination level	Public

This document is Deliverable 2.2 of Work Package 2 of the e-SIDES project on Ethical and Societal Implications of Data Science. e-SIDES is an EU funded Coordination and Support Action (CSA) that complements Research and Innovation Actions (RIAs) on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies and providing a broad bases and wider context to validate privacy-preserving technologies. All interested stakeholders are invited to look for further information about the e-SIDES results and initiatives at www.e-sides.eu.

Grant Agreement number: 731873

Grant Agreement number: 731873

Executive Summary

The main aim of this document is to identify and analyse the most relevant ethical, legal, societal and economic issues implicated by the development of big data technologies. With this purpose in mind, each distinctive perspective approaches the technological innovation brought about by big data technologies from a different angle.

First, the **ethical perspective** contains a comprehensive review of different ethical outlooks: moral philosophy, philosophy of technology and biomedical ethics which provide the guidelines for developing a list of values that are useful to shape an ethical perspective on big data technologies for all stakeholders. The ethical issues mapped particularly concern these values to the extent they are under pressure by the developments in big data technologies. The selection was primarily guided by the views on technology development from a virtue ethics perspective. The ethical issues identified are: *human welfare, autonomy, non-maleficence, justice (including equality, non-discrimination, digital inclusion), accountability (including transparency), trustworthiness (including honesty and underpinning also security), privacy, dignity, solidarity and environmental welfare*.

Second, the **legal perspective** focuses on the lists of human rights derived from the European Convention on Human Rights (ECHR) and the EU Charter of Fundamental Rights (the EU Charter), which together constitute the main legal framework for the EU in the field of human rights. The rights of particular relevance in the context of big data technologies are the rights to private and family life, personal data protection, freedom of expression and information, freedom of assembly and association, non-discrimination, fair trial and consumer protection. By analysing the normative scope of each of these human rights, looking at both legislation and case law of the European courts and the way in which big data technologies challenge different aspects of each human right at stake, the legal part distils the list of the most relevant issues at the nexus of big data technologies and human rights in the EU. The legal issues identified are: *lack of transparency, vagueness of the concept of harm, accountability, proportionality, establishing a regulatory framework and the role of private actors in applying fundamental rights*.

Third, the **societal perspective** makes use of the extensive literature on Societal Impact Assessments (SIA). The analysis of literature was combined with a review of research project propositions and complemented by discussions at two workshops. Societal impact is very generally understood as changes to one or more of a number of elements of social life: people's way of life, their culture, their community, their political systems, their environment, their health and well-being, their personal and property and their fears and aspirations. The societal issues are mapped by examining different actors and distinctions between these actors, by examining the relationship between data subjects and data controllers and processors, and by examining the risk and impact of potential abuses of big data technologies. On top of the SIA approach, a survey of literature on societal issues in the context of big data technologies identified data culture, data quality, analytics methodology and visualisation as related aspects, essential to understand societal issues and to develop means to address them. The societal issues identified are: *unequal access, normalisation, discrimination, dependency, intrusiveness, non-transparency and abusiveness.*

Fourth, the **economic perspective** mainly builds on the societal perspective, as the societal perspective already includes business-to-business and business-to-consumer relations. Societal issues may affect community capital, which may include human capital, social capital, political capital and cultural capital.

Natural and physical capital are outside the scope of this deliverable. Due to this close relationship between the societal and the economic perspective, many of the societal issues also include economic aspects and, as such, societal and economic issues cannot always be clearly distinguished. Therefore, the starting point for listing the economic issues are the societal issues derived from the SIA analysis, with an emphasis on economic aspects. There are no economic issues that are not societal issues at the same time. The economic issues identified are: *unequal access (including the shortage of a skilled workforce and the creation of a new digital divide), normalisation, discrimination, dependency, intrusiveness, non-transparency and abusiveness*.

Observing the four lists of issues identified, the following conclusions can be drawn:

- Although there is some **overlap in issues** from the different perspectives, this does not mean that the overlapping issues are the same from each perspective – each perspective simply shows **different aspects of each issue**.
- The list of issues identified is **very extensive**, but **not exhaustive**. The rapid changes in big data technologies call for **periodic updates** of identification of issues.
- The issues identified are **hard to prioritize**, as this may be **context-dependent** and many issues are **interconnected**.
- The issues identified should not only or merely be regarded as problems to be solved, but rather as providing the **goals to strive for**. An attitude of **continuous attention is required** for these issues.

These conclusions call for further work. The inventory in this deliverable may require periodic updates after some time. Furthermore, balancing and prioritizing the issues identified is hard *in abstracto* and may, therefore, call for more detailed, context-specific approaches. Finally, because many of these issues cannot be solved once and forever, an attitude of continuous attention for these issues is called for.

Grant Agreement number: 731873

Contents

Grant Agreement number: 731873

Grant Agreement number: 731873

1. Introduction

1.1. Background

This report is Deliverable 2.2 of the e-SIDES project. In this project the ethical, legal, societal and economic implications of data sciences (particularly big data and big data technologies) are examined in order to complement the research on privacy-preserving big data technologies. The first step in this project is to identify the most important ethical, legal, societal and economic issues related to data science and big data, which is the aim of this deliverable.

The differences between these four perspectives are explained in deliverable D2.1 of this project. The ethical issues and the legal issues are typically closely related. For instance, some ethical principles and values are codified in EU legislation, which may provide more concrete requirements for the design of privacy-preserving big data technologies. Also, some ethical issues that are identified and validated in e-SIDES may feed recommendations on changing existing EU legislation. However, the scope of legal issues is restricted to the ethical issues of big data technologies. Legal issues that are not (also) ethical issues, such as difficulties regarding the enforcement of particular legislation, are beyond the scope of e-SIDES. All ethical and legal issues will primarily be mapped from a European perspective.

Whereas ethical issues focus on whether something is right or wrong, societal issues focus on how society is affected by something. Ethical and societal issues are usually related to each other. There are only a few examples that can be fully put under only one of the two categories. When addressing societal issues, economic issues are closely related to these, as big data technologies implicate economic issues that may have positive and negative societal consequences. Examples for such issues in the big data context are productivity growth and downsizing of work forces. Further issues that are taken into account from a societal perspective – and probably also from an ethical perspective – include trust, discrimination, inequality of access, exploitation and manipulation. With respect to societal and economic issues, e-SIDES focuses on the European multicultural and multinational landscape.

Hence, in summary, this deliverable maps issues related to big data and big data technologies from four different perspectives, which are also shown in Figure 1:

- The ethical perspective
- The legal perspective
- The societal perspective
- The economic perspective

Grant Agreement number: 731873

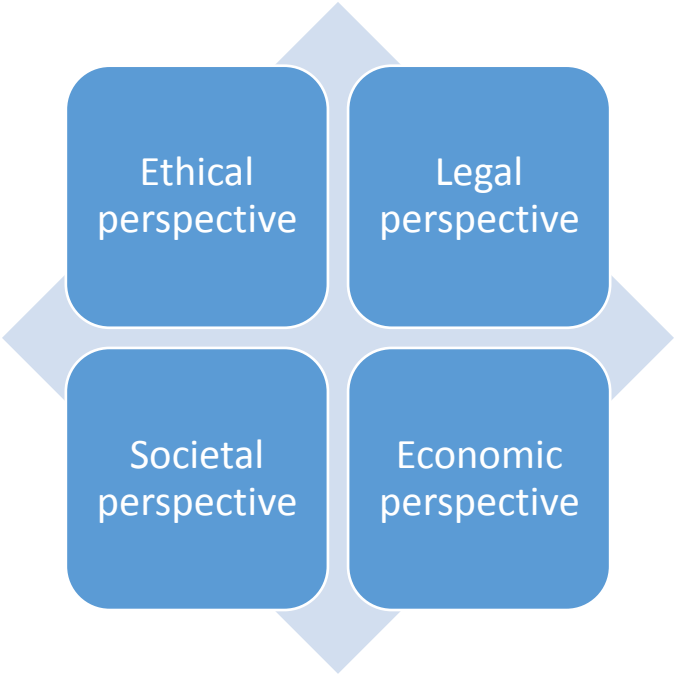

Figure 1 The four perspectives taken on big data technologies

The selection of issues in this deliverable is not exhaustive, but it represents a rich set of relevant perspectives on big data-technologies. This selection is aimed to form useful stepping stones for the assessment of privacy-preserving big data technologies (both existing and under development) and also for the development of design requirements for these technologies in the next deliverables of this project. Such design requirements are intended to prevent or remedy undesirable side-effects of big data and big data technologies, such as privacy violations, discrimination, unfair treatment, social disparity and other forms of injustice.

The remainder of this chapter is as follows. In the next section, Section 1.2, the methodology to create the lists of ethical, legal, societal and economic issues is discussed. In Section 1.3 the chapter structure of this deliverable is explained.

1.2. Methodology

In this section, the methodology used to identify issues is explained. This section starts with describing the general approach (subsection 1.2.1), next discusses the methods used (subsection 1.2.2) and then explains the particularities for each perspective dealt with in this deliverable (subsection 1.2.3).

1.2.1. General approach

The goal of this deliverable is to create lists of ethical, legal, societal and economic issues relevant in the context of big data technologies. To create these lists, several approaches were used, both for making inventories and for validating the results. In this subsection we describe the general approaches used for

Grant Agreement number: 731873

creating the lists of issues. In the next subsection we describe the specific approaches used for the different perspectives, as each of the disciplines investigated has its own methods.

When creating the lists of issues, important considerations are completeness and overlap. Starting with the first consideration, completeness, we obviously want to create lists of issues that are as complete as possible. That is why a comprehensive, multi-method approach is chosen to map the issues. However, although a complete, exhaustive list of issues is the ideal goal of this deliverable, it is important to note that it is impossible to provide such a complete, exhaustive list for two reasons. First, there exists no theoretical framework that is a closed system in which it is possible to take an exhaustive approach. Second, even if a closed system were available, it would be impossible to take an exhaustive approach because the big data technologies and applications are rapidly changing all the time.[1] Nevertheless, the comprehensive approach taken in this deliverable makes it very likely that the most important issues in each of the disciplines investigated are actually mapped.

The other consideration, overlap, is also important. Ideally, we would like to create a list of issues that do not overlap. However, also in this case it is important to note that a list of non-overlapping issues is not realistic. Such a non-overlapping categorisation does not exist in any of the disciplines that represent the different perspectives in this project. Hence, there will be some overlap in the issues in two ways. The first is overlap in issues in the different perspectives. For instance, privacy is both an ethical and a legal value. Another example is transparency, which may be important from a societal perspective (for social acceptance), but may also be an economic issue (reputation or trade secrets). Second, there may be some overlap in the issues themselves. For instance, privacy (an issue with regard to preventing unwanted disclosure of sensitive personal information) may overlap with security (as data breaches may yield such unwanted disclosure). Yet, privacy and security are different issues. Another example may be automated decision-making on the basis of big data analytics, in which the autonomy of data subjects may be infringed but also transparency about these processes may be lacking. Again, autonomy and transparency are different issues. Finally, it should also be mentioned that, as the four perspectives that are combined in this project are inherently connected, it cannot be avoided that some of the issues are also discerned by more than one perspective.

1.2.2. Methods

The results in this deliverable are mainly based on desk research. The stocktaking of issues was based, in the first place, on a thorough **review of related scientific and practitioner literature**. For this, literature was collected from the domain of the respective perspectives (i.e., the ethical, legal, societal and economic perspectives) and from the domain of big data and data sciences, particularly literature on privacy preserving big data technologies. Literature that addressed any ethical, legal, societal and economic issues relevant in the context of big data technologies was selected for further processing. Particularly literature with lists of issues was focused on. Furthermore, a systematic review of past and

[1] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, p. 120.

Grant Agreement number: 731873

ongoing (European) projects analysing the impacts of emerging technologies was part of this desk research.

The issues resulting from this literature review was supplemented with **expert knowledge**, from both the researchers in this project and external experts. The expert knowledge was used to assess the completeness and exhaustiveness of the collected results (see above). Where apparent gaps appeared, expert knowledge was used to add further issues to the lists and/or further qualify the issues that were already on the lists. Furthermore, expert knowledge was used to categorise the results into clear, well-defined categories that avoid overlapping (to the extent possible, see above). Experts were consulted within the professional networks of the researchers involved in this project, but also at expert forums.

After the literature review and the use of expert knowledge, as a third step, two **workshops** were organised to further collect any missing issues, to obtain additional in-depth knowledge on the issues mapped, and, to validate the then-preliminary results. The ultimate objective of the workshops (both titled "Societal and Ethical Challenges in the Era of Big Data: Exploring the emerging issues and opportunities of big data management and analytics") was to discuss the main ethical, legal, societal and economic challenges emerging from the adoption of big data technologies thus helping the researchers perform responsible research while pursuing innovation. The workshop format was designed to allow for the maximum participation and interaction of the attendees, with a live survey to gather and discuss opinions. The questions raised in both workshops can be found in Appendix A and B of this deliverable. In particular, the workshop focused on to what extent the participants considered these challenges relevant in the big data arena and thus to be taken into account by the community.

The first workshop was held at the CEPE/Ethicomp Conference 2017 "Values in Emerging Science and Technology" at Turin University on June 7th to collect views and opinions on the importance of these issues and on how the researchers and technologies developers should take them into account. In the discussion, the insights focused in particular on how big data technologies could lead to discriminatory treatment of certain groups. Not only can such technologies amplify the already existing biases and divides in the society, big data technology could also potentially have the transforming effect of creating different perceptions and 'new normals' in society. Furthermore, the workshop discussion focused on the impact of the provisions of the General Data Protection Regulation (GDPR) on the development of big data in the EU, the margin left for the member states' regulation and possible approaches to interpreting the GDPR in this context.

The second workshop was held at the ICE/IEEE Conference 2017 in Madeira, which aimed at bringing together high-level research and business community around the topic: "Engineering, Technology & Innovation Management Beyond 2020: New Challenges, New Approaches". The initial set of questions, in particular the formulation of the statements, were revised and improved on the basis of the feedback received and on the expert opinion of the project team. The results of the workshop voting sessions are presented and analyzed in the next chapters. For an overview, see Appendix A and B of this deliverable.

1.2.3. Specific approaches for the different perspectives

The differences in the four perspectives used in this project are described in Deliverable 2.1. Because the four perspectives represent different disciplines, they each call for a more specific approach. In this subsection the specific approaches for the different perspectives are explained in more detail.

Ethical issues approach

When mapping ethical issues, the first question obviously is what constitutes an ethical issue. For this purpose, we distinguish three types of ethical issues:

1. Ethical issues may occur when moral principles (taken broadly) are violated.
2. Ethical issues may occur in situations in which moral principles (taken broadly), or actions-directives resulting from the application of moral principles to particular situations,[2] conflict with each other (i.e., *moral conflicts*).
3. Ethical issues may occur when new (types of) problems arise for which no moral principles (taken broadly) exist, or when it is not clear which principles to apply to particular cases.

Now, this leads to the question what moral principles are. In order to explain this, we start with the differences between ideals and values, (moral) principles and (moral) rules. Ideals and values are more abstract and general than principles; principles, in turn, are more abstract and general than rules.[3] Starting at the most concrete, specific level, *(moral) rules,*[4] in their most common form, are prescriptive norms of conduct: they determine how one ought to behave, e.g., do not smoke, pay your bills, do not steal.[5] *Principles* may be considered to be norms that prescribe that a thing be realised to the highest degree *possible* in a particular situation.[6] Principles can be fulfilled to a certain degree, whereas rules are either fulfilled or not.[7] For instance, a rule about health care may specify to what extent care has to be taken, while a principle about health care requires that the best care possible be taken, which makes it open-ended. A principle may be considered a norm prescribing how to go about the realisation of ideals or values in a particular situation. In other words, *ideals* or *values* may be considered the desirable state of

[2] See, for instance, Ross (1930), who argues that moral conflicts may sometimes be caused by a single moral principle rather than by two conflicting moral principles. For instance, when two people are starving but there is sufficient food to save only one of them, the principle of beneficence states that one should be saved, instead of both being allowed to die. This (single) moral principle does not solve the problem, however, since it does not tell us to whom the food should be given. Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press.

[3] Note that this hierarchy may not always be as obvious as suggested here, but for the purposes of this deliverable these descriptions should be sufficient.

[4] Note that there is a difference between *legal rules* and *moral rules*. Moral rules will be dealt with in the ethical issues, legal rules will be dealt with in the legal issues.

[5] Rules are not always simply prescriptive. See, for instance, Hart, H.L.A. (1994) *The Concept of Law*, Oxford: Clarendon Press, p. 91, who draws a distinction between primary rules, i.e., rules of obligation, and secondary rules, i.e., rules about the primary rules, such as rules about applying and changing rules.

[6] See also Alexy, R. (1985) Rechtsregeln und Rechtsprinzipien, *Archiv für Rechts- und Sozialphilosophie*, Beiheft 25, p. 13-29.

[7] See also Dworkin, R. (1978) *Taking rights seriously*, Cambridge: Harvard University Press, p. 24.

affairs that one aims to realise, and principles are the norms prescribing how to achieve this state of affairs. For instance, the principle of justice[8] prescribes how to realise an ideal of a just society. How this desirable state of affairs (e.g., a just society) looks like, may not be known in detail. For the purpose of this deliverable, we focus on moral principles.

In the 21st century, existing moral principles may need new and explicit adaptation to our emerging global technomoral environment.[9] While the core meaning of moral principles may be fixed over time and over generations of people, the concrete meaning may be determined by the distinctive shape of a specific moral context. Especially in the domain of big data and data science, the context is one of increasingly rapid, transformative, global, unpredictable and interdependent technosocial change.[10] As technosocial conditions changes over time, moral principles may have to evolve with them.

In the realm of normative ethics, three main theories are generally distinguished: utilitarianisms, Kantianism, and virtue ethics. All these theories address the same questions: What should we do and how should we act? What are the principles underpinning our actions? Or in short, what is morality? This may translate in more practical questions such as: do we have a moral duty to tell the truth? And if so, why? Although these theories might address by and large the same question, the way in which they answer it, differs.

In Kantianism or deontology, morality is based on the universal law of rationality, also referred to as the categorical imperative[11]. The most important moral criterion for the Categorical Imperative is its universability. Why should I not lie? Because, if everyone would lie, there would simply be no way to distinguish true from false. It would become impossible to make rational decisions and this in the end would diminish human dignity. This is unattainable and hence not in line with the law of rationality. Universal rational consistency is key to the Kantian approach.

For utilitarianism or consequentialism, morality is about maximising the amount of good things such as pleasure and happiness and minimising the reverse. The goal of morality in utilitarian theories is happiness or well-being. For utilitarianists lying may therefore well be the right action if it adds to the overall well-being and happiness of most people.

Both Utilitarianism and Kantianism are characterized by rather fixed rules and principles and they both approach morality as impartial. In utilitarian theory, there is no difference between my own happiness and that of someone else.[12] Morality revolves around maximising happiness at large and not my happiness specifically. For Kantianism, empirical considerations (such as introspection, psychology, biology,..) should not play a role in deciding what a moral good is. What is morally good cannot be grounded upon a non-moral good.

---

[8] Note that there may be different interpretations of the principle of justice.
[9] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, p. 119.
[10] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, p. 119.
[11] Kant, E. (2012) Groundwork of the Metaphysics of Morals (1785). Revised Edition. Ed. M. Gregor and J. Timmermann. Cambridge: Cambridge University Press.
[12] Crisp, R and M. Slote (1997) Virtue Ethics. New York, Oxford University Press, p.1

Grant Agreement number: 731873

Both theories have been criticized. Utilitarianism has been criticized for the fact that it decouples the moral worth of acts from the moral worth of persons, consequently sometimes legitimizing the well-being of a minority being sacrificed for a greater overall happiness.[13] The Kantian approach has been disapproved because it appreciates rational consistency more than human relationships of care.[14]

The rise of modern virtue ethics can be seen as a reaction to both Kantianism and utilitarianism. Unlike these normative theories, virtue ethics does not have a supreme principle of morality. An action is judged as being morally right when it can be identified as an action that a virtuous person would undertake. "Virtues are desirable qualities of persons that predispose them to act in a certain manner".[15] It is assumed that acting virtuously will lead to human flourishing. But what this flourishing entails depends on the situation and can change over time. In virtue ethics, it is therefore not enough to merely apply moral principles such as those central to Kantian or utilitarian ethics in order to decide how to act. What a particular virtue consists of depends on the context and on that person. In virtue ethics, both rationality as well as more contextual and relational considerations must therefore be taken into account.

As it is impossible to determine the right action a priori – so without taking the contextual elements into account – virtue ethics emphasizes the importance of practice. People should have the opportunity to explore and develop their virtues. They need to learn how to distinguish what the important moral issues are in a particular situation and how to creatively address those. For example, just as in Kantianism, a virtue ethicist will tell you that lying is wrong. However, not because it is against the categorical imperative and therefore always wrong, but because dishonesty leads you further away from being a virtuous person. Lying is a so-called vice. However, this does not entail that you therefore should always tell the truth. A truly virtuous person in addition also knows when, to whom and where what kind of information should be shared.[16]

One can argue that contrary to Kantianism and utilitarianism, virtue ethics is less systematic and lacks clear criteria on which you can base your actions. However, the more flexible theoretical structure of virtue ethics may also better accommodate the exploration of what ethical life in the big data era may look like. It does not conceive human beings as merely rational agents, but also pays attention to their emotional disposition, their relations and the social context in which they operate. Because virtue ethics values practical wisdom above fixed rules it is "ideally suited for adaptation to the open-ended and varied encounters with particular technologies that will shape the human condition in this and coming centuries"[17]. Moreover, because of its focus on practical wisdom, a virtue ethics approach is well suited to include the insights of stakeholders gained at workshops and other related meetings in the analysis. Virtue theory as a model has been adopted in several domains including bioethics, media ethics, and

[13] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, p.23.
[14] idem.
[15] Mazur, T. 1993 Lying. Ethics - V. 6, N. 1 Fall
[16] idem, p.19
[17] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, p.33

Grant Agreement number: 731873

business ethics. Recently also scholars in ethics and philosophy of technology have adopted this approach.[18]

In the light of these considerations, we choose virtue ethics, rather than Kantianism or utilitarianism as the main approach in mapping ethical issues. By taking virtue ethics as our main approach, we are challenged to reflect upon the question what a) human flourishing in the big data era comes down to; b) what the necessary conditions are to ensure this flourishing and c) what people need in order to bring these virtues in practice. The latter is closely connected to the privacy-preserving technologies that are central to this project. What are the ethical design requirements for these technologies to make sure that citizens can develop and practice their virtues? In other words, in which way can privacy-preserving technologies facilitate people becoming virtuous persons? The goal is to develop a list of ethical issues or challenges that are central to the use of big data and related virtues that actors such as citizens/users/companies should put into practice in order to ensure human flourishing in the big data era.

Legal issues approach

The process of selecting legal issues relevant in the context of big data technologies is determined by the stipulation that such issues should be connected to the ethical considerations. This further focuses the scope of the legal perspective taken into consideration to the sphere of human rights. Additionally, as the focus of this deliverable is on the European legal framework, the legal sources for the protection of human rights taken into account are limited to the main European legal instruments in this regard: the European Convention of Human Rights (ECHR)[19] and the European Charter of Fundamental Rights (EU Charter).[20] The determination of the relevant issues is based on the research of both the relevant literature in this field as well as extensive research of the case law of the European Court of Human Rights (ECtHR) as well as the Court of Justice of the EU (CJEU).

In order to determine a list of issues relevant in this context a three-step methodology was applied. First, a general framework for the protection of human rights in the EU is presented with the purpose in mind to determine the specific institutional and legislative environment that governs the protection of fundamental rights. This preliminary outline allows demonstrating the perspectives for and limitations of using these human rights instruments in the context of new technological realms, particularly in the field of data science and big data. Moreover, the overview of the EU legal framework for the protection of human rights gives a reader non-familiar with the legal domain an opportunity to understand the main rules and processes governing this field and thus have a better understanding of the subsequent sections of Chapter 3.

Second, the list of human rights relevant in the context of the big data technologies is selected. At the outset, three distinctive stages implicated in the use of big data are discussed: data gathering, data processing and application of the derived knowledge. This distinction is useful as different stages might

[18] Idem. See also Sloot, van der, B. (2017) Privacy as Virtue. Moving Beyond the Individual in the Age of Big Data.
[19] European Convention for the Protection of Human Rights and Fundamental Freedoms, 1953.
[20] European Union, Charter of Fundamental Rights of the European Union [2000] OJ C364/01 and [2010] OJ C83/389.

Grant Agreement number: 731873

turn out to be challenging for different human rights. For example: the stage of data gathering typically triggers the interests behind the right to data protection but might be completely irrelevant from the perspective of the right to non-discrimination. Subsequently, based on the catalogue of human rights as set forth in the ECHR and the EU Charter and having in mind the specific characteristics of big data technologies, the list of the most relevant rights is selected. The list is not comprehensive, as the variety of the specific contexts in which big data technologies are applied as well as the speed of developments and level of advancement of these technologies allow only for addressing the human rights which are most significantly and directly impacted. Further, the list of relevant fundamental rights established in this manner is submitted for the twofold analysis. First, the general substantive scope of each right is discussed, with the particular focus on the features that are of relevance in the context of big data. Such discussion is driven mainly by the case law analysis of the two courts and allows for determining the specific normative framework of each human right listed. While the ethical part of this deliverable focuses on the values which also underpin the human rights placed on the list, the aim of the legal analysis is to outline the normative framework underlying human rights law that encapsulates the moral values and giving rise to certain specific rights and obligations. Second, following determining of the normative framework, each right is specifically discussed in the context of challenges brought about by big data technologies. Thus, this second-stage discussion allows illuminating which aspects of the normative framework are put under pressure by the big data technologies.

Finally, following the review of different ways in which different human rights are challenged, the list of the most relevant legal issues in this context is distilled. Pursuant to this list, the legal issue is an issue which is of relevance in the context of normative framework of one or more human rights and proves to be particularly challenged by the big data technologies. The legal issues concern the legal framework: they focus on the way in which legal norms aimed at ensuring the protection of human rights are rendered futile or difficult to apply. In particular, the legal issues are divided into two categories: (1) the issues concerning challenges to the ensuring the protection of particular human rights and (2) the issues concerning the functioning of the normative framework of fundamental rights as a whole. The final list is not comprehensive as it is virtually impossible to address all the possible ways in which big data technologies challenge the current framework of human rights. However, as the list of legal issues contains the broad categories, it allows for addressing the myriad of different specific matters which give a good overview of the spectrum of legal challenges at the nexus of big data technologies and human rights.

Grant Agreement number: 731873

Societal and economic issues approach

The identification of societal and economic issues is also mainly based on desk research. On the one hand we make use of the extensive literature on "social impact assessment" (SIA) that has been published since SIA was developed as a spin-off and extension of "environmental impact assessments" since the 1980s.[21] This has been complemented by literature on approaches for integrated impact assessments, e.g. under the umbrella term "Responsible research and innovation" (RRI).[22] Apart from these more general sources we extensively scan mostly EU-funded projects that (also) aim to assess (social) impacts of emerging technologies in the field of big data technologies.[23] The analysis of academic literature and the review of project positions were complemented by discussions at two workshops and a thorough investigation of studies and reports with economic focus.

According to the International Association for Impact Assessment (IAIA), a SIA includes the "process of identifying and managing the social issues of project development, and includes the effective engagement of affected communities in participatory processes of identification, assessment and management of social impacts."[24] In this respect, a SIA is a practically orientated concept.

Very generally the IAIA defines societal impacts as changes to one or more of a number of elements of social life: people's way of life, their culture, their community, their political systems, their environment, their health and wellbeing, their personal and property and their fears and aspirations.[25] Moreover the IAIA distinguishes different forms of community capital that can be affected. For the purpose of this project, we will not consider natural and physical capital[26] and only partly consider financial capital. What has to be considered also for big data technologies, are the impacts on the following:[27]

- *Human capital* refers to skills and abilities of people to develop and enhance their resources and to access outside resources and bodies of knowledge
- *Social capital* reflects the connections among people and organizations or the social "glue" to make things happen. This comprises social networks and trust but also social rules, norms and obligations and reciprocity arrangements

[21] For an overview see Bianca Dendena and Stefano Corsi, "The Environmental and Social Impact Assessment: A further step towards an integrated assessment process," *Journal of Cleaner Production* 108 (2015)

[22] For an overview see Richard Owen, John Bessant and Maggy Heintz, eds., *Responsible Innovation* (Chichester, UK: John Wiley & Sons, Ltd, 2013) and E.-M. Forsberg et al., "Assessments of emerging science and technologies: Mapping the landscape," *Science and Public Policy* 41, no. 3 (2014)

[23] The projects included inter alia ASSERT, BIG, BYTE, DESSI, SysSec, CRISP, Coco Cloud, CAPITAL, RESPECT, Socialising Big Data (SBD), CLARUS, DwB, ENFORCE, SURVEILLE, EuroPriSe, PIAF, CONSENT, SAPIENT. Information about these projects can be found in the CORDIS database: http://cordis.europa.eu/projects/home_en.html

[24] Frank Vanclay et al., "Social impact assessment: Guidance for assessing and managing the social impacts of projects," (International Association for Impact Assesment (IAIA), 2015), https://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf (accessed August 11, 2017)

[25] Frank Vanclay, "International Principles For Social Impact Assessment," *Impact Assessment and Project Appraisal* 21, no. 1 (2003): 8

[26] Natural capital mainly inlcudes stocks and flows of environmental assets, such as food and forestal resources, mineral reserves, water, soil, air etc. Physical capital comprises the stock of equipment, physical plants, infrastrutcure an other productive resources.

[27] Vanclay et al., "Social impact assessment", 13 and Mary Emery and Cornelia Flora, "Spiraling-Up: Mapping Community Transformation with Community Capitals Framework," *Community Development* 37, no. 1 (2006): 20f

- *Political capital* stands for the access to power, organizations, connection to resources and power brokers. It refers to the existence and effective functioning of governance mechanisms, i.e. standards, rules, regulations and their enforcement
- *Cultural capital* finally reflects the way people define their place within the world and how they act within it. This includes language and traditions and their function for social inclusion and development.

Since many of the elements also include economic aspects societal and economic issues cannot always be clearly distinguished. It is quite evident that there is also a significant overlap with ethical and legal issues: in particular the respect for human rights is a crucial element in most of the community capital categories.[28]

We use these rather broad categories as a benchmark to make sure that the much more specific issues found in the literature about the effects of big data technologies fully cover the spectrum of possible societal and economic impacts.

1.3. Structure

This deliverable is structured as follows. The next four chapters focus on each of the four perspectives in this project respectively. Hence, Chapter 2 identifies ethical issues, Chapter 3 identifies legal issues, Chapter 4 identifies societal issues and Chapter 5 identifies economic issues. Although each chapter has a slightly different set-up due to the specific nature of each of the perspectives, they are similarly structured in the sense that they all start with a table of issues and respective explanations. Next, it is explained how these lists were compiled, by identifying sources of (lists of) issues, then analyse the issues identified and distil from this (through processes of weighing, prioritizing and aggregating) a final list of issues for the e-SIDES project. The final chapter, Chapter 6, provides conclusions.

[28] Deanna Kemp and Frank Vanclay, "Human rights and impact assessment: Clarifying the connections in practice," *Impact Assessment and Project Appraisal* 31, no. 2 (2013) and Vanclay et al., "Social impact assessment"

2. Ethical perspective

e-SIDES: values for big data technologies	*Issues putting pressure upon values in the context of big data technologies*
Human welfare	Discrimination of humans by big data-mediated prejudice can occur. Detrimental implications can emerge in the contexts of employment, schooling or travelling by various forms of big data-mediated unfair treatment of citizens.
Autonomy	Big data-driven profiling practices can limit free will, free choice and be manipulative in raising awareness about, for instance, news, culture, politics and consumption.
Non-maleficence	Non-transparent data reuse in the world of big data are vast and could have diverse detrimental effects for citizens. This puts non-maleficence as a value under pressure.
Justice **(incl. equality, non-discrimination, digital inclusion)**	Systematic unfairness can emerge, for instance, by generating false positives during preventative law enforcement practices or false negatives during biometric identification processes. (Such instances put constant pressure on the value of justice.)
Accountability **(incl. transparency)**	For instance, in the healthcare domain patients or in the marketing domain consumers often do not know what it means and who to turn to when their data is shared via surveys for research and marketing purposes.
Trustworthiness **(including honesty and underpinning also security)**	Citizens often do not know how to tackle a big data-based calculation about them or how to refute their digital profile, in case there are falsely accused, e.g.: false negatives during biometric identification, false positives during profiling practices. Their trust is then undermined. The technology operators trust at the same time lies too much in the system.
Privacy	Simply the myriad of correlations between personal data in big data schemes allows for easy identifiability, this can lead to many instances for privacy intrusion.
Dignity	For instance, when revealing too much about a user, principles of data minimization and design requirements of encryption appear to be insufficient. Adverse consequences of algorithmic profiling, such as discrimination or stigmatization also demonstrate that dignity is fragile in many contexts of big data.
Solidarity	Big data-based calculations in which commercial interests are prioritized rather than non-profit-led interests, are examples of situations in which solidarity is under pressure. For instance, immigrants are screened by big data-based technologies, they may not have the legal position to defend themselves from potential false accusations resulting from digital profiling which can be seen as a non-solidary treatment.
Environmental welfare	Big data has rather indirect effects on the environment. But for instance, lithium mining for batteries is such. (But extending the life-expectancy of batteries and, for instance, using more sun-energy for longer-lasting batteries could be helpful.)

Table 1 Overview of the identified ethical issues

Grant Agreement number: 731873

2.1. Ethical issues in big data technologies

Big data is generally considered the new gold. Richards and King have compared its implications with that of the "Industrial Revolution".[29] Although the variety of big data benefits can surmount our imagination, through the exponential growth in digital devices and their computational capabilities big data puts constant pressure upon the boundaries of what can or cannot be seen as acceptable in a society from an ethical perspective. Literature is steadily expanding concerning the latter dilemmas.[30] Although the intentions behind big data technologies are noble, a variety of issues emerges purely from their immense and exponentially growing capacities that are on the one hand facilitated by their design and on the other further cultivated by their use. In this chapter, we will therefore map recent literature on ethics and big data technologies, especially literature that contains lists of issues, and make a selection of issues, which seem useful to account for in relation to big data technologies in our society.[31]

Ethical codes that entirely match or guide the path of growth in big data technologies do not exist. Nevertheless, moral philosophy, philosophy of technology and even biomedical ethics can provide a set of values that are useful to shape an ethical perspective on big data technologies for all stakeholders including designers, policy makers, users and, of course, data scientists. Furthermore, we acknowledge that when choosing these values our motivation has been influenced by views on technology development from the virtue ethics perspectives. Although mapping currently existing deeply-rooted philosophical discussions in utilitarian ethics[32] (J. Mill, 1886) and deontological ethics[33] (Bentham, 1789; Kant, 1780) could be an interesting academic exercise, we use an applied ethics (and more precisely applied virtue ethics)[34] approach to map ethical issues. The choice for virtue ethics is motivated in the methodology section (particularly Subsection 1.2.3). Furthermore, virtue ethics provides the pillars for the

[29] Richards, N. M., & King, J. H. (May, 19, 2014). Big Data Ethics. *Wake Forest Law Review*, 2014

[30] Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good Frameworks for Engagement* (pp. 44–75). Cambrige University Press.; boyd, danah and Crawford, Kate, Six Provocations for Big Data (September 21, 2011). A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. Available at SSRN: https://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431; Engin, I., & Ruppert, E (2015). *Being Digital Citizens.* London: Rowman & Littlefield International; Hasselbalch, G., & Tranberg, P. (2016). *DATA ETHICS - The New Competitive Advantage*. PubliShare.; Modderkolk, H. (2015). Met big data alleen ga je echt geen aanslagen voorkomen. *Nos.nl*. Retrieved from http://www.volkskrant.nl/buitenland/met-big-data-alleen-ga-je-echt-geen-aanslagen-voorkomen~a4192661/?hash=642ef3fff40bd5faffc383042424afe251927b52; Strandburg, K. (2014). Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good : Frameworks for Engagement*. Cambridge University Press.; Zook, M., Barocas, S., Crawford, K., Keller, E., Goodman, A., Hollander, R., … Pasquale, F. (2017). Ten simple rules for responsible big data research. *Computational Biology*, *13*(3), 1–10.

[31] In some cases we list principles or values *under pressure*, which are a specific type of ethical issue (see Section 1.2).

[32] Utilitarianism - moral theory that guides choices by what is desirable or good as an end to achieve.

[33] Deontology - moral theory about which choices are required or forbidden or what we ought to do. In other words deontology is a study of (external or internal) moral obligations.

[34] Moral theory on what kinds of persons we should be, how we can lead a 'good life' that is most worthy of us.

development of fundamental rights schemes in Europe and therefore also constitutes fruitful grounds for positioning the legal section in this deliverable (see Chapter 3).

2.2. Lists of values (virtues) from philosophy of technology

2.2.1. List of 'technomoral virtues'

A first approach to map ethical issues is to apply the moral virtues that have existed for centuries to big data technologies. However, since these virtues were not developed in the big data era, they may need adjustment and explicit adaptation. Shannon Vallor is a pioneer philosopher in investigating current intricacies at the cross-roads of big data technologies and virtue ethics. She points out the acute need for exploring what ethical values and norms are about to erode in our big data-driven society. Her analysis is based on ancient Greek philosophers like Aristotle, medieval Christian philosophers like St Thomas of Aquinas and also Eastern philosophies like Confucianism or Buddhism. Inspired by perspectives on virtues which stem from long before the existence of our current big data technologies, she argues that: "we need to cultivate in ourselves, collectively, a special kind of moral character, one that expresses what I will call the technomoral virtues." In line with this train of thought, she defines a list of values[35] that in her view are such technomoral virtues that deserve upholding. These virtues may help us identifying ethical issues. The list of technomoral virtues is as follows:

Honesty: "Respecting truth"

"Flourishing in interactions with other people" which Vallor argues as being the primary task of ethics, would be "impossible without the general expectation of honesty".

Self-control: "Becoming the author of our desires"

Vallor explains self-control as a person's "ability to align one's desire with the good".

Humility: "Knowing what we do not know"

Vallor describes this value as something not always having been as stably acknowledged as the above-mentioned three other values. Yet, its importance should be acknowledged especially in the era of big data. Vallor describes humility as a technomoral virtue as: the *"recognition of the real limits of our technosocial knowledge and ability, reverence and power at the universe's power to surprise and confound us, and the renunciation of the blind faith that new technologies inevitably lead to human mastery and control of our environments".*

Justice: "Upholding rightness"

Vallor argues about justice as being the broadest among values and based on Aristotelian and Confucian understandings she explains this as the "just treatment of others" that can be understood under the broader notion of human benevolence. In Buddhism she explains that justice means the "unconditional concern for the welfare and dignity of all creatures". Vallor explains, for instance, that among others

[35] This list is from Shannon Vallor (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, pp. 120-121.

growing social inequalities are currently fuelled by "data-mining, pervasive digital surveillance and algorithmic profiling, robotics, drones" and technomoral justice is necessary to restore or remedy such inequalities.

Courage: "Intelligent Fear and Hope"

Vallor interprets acting courageously as being able to "intelligently balance measured and justified fears with measured and justified confidence and hope". Courage is fundamentally different from self-control in the sense that it invites for taking risks and sacrifices for the purpose of the good. In the era of emerging technologies Vallor calls for actively embracing this value by cultivating a "reliable disposition toward intelligent fear and hope with respect to the moral and material dangers and opportunities presented by emerging technologies". This is something also highly relevant for analysing privacy-preserving technologies.

Empathy: "Compassionate Concern for others"

Vallor explains empathy as a technomoral virtue as being "a cultivated openness to being morally moved to caring action by the emotions of the other members of our technosocial world." This could mean for cultivating compassion both for others' joy as well as pain. Vallor calls for embracing this value, because the amount of joyful and painful events that are mediated via digital technologies is unseen, therefore we can too often be called for empathic action. Nevertheless, we need to adapt to these technological changes while adapting our receptors of empathy to these changes. This seems crucial in order not to become narcissists.

Care: "Loving serving to others"

Technomoral care Vallor defines as the "skilful, attentive, responsible and emotionally responsive disposition to personally meet the needs of those with whom we share our technosocial environment." She points out a danger of emerging technologies for the cultivation of human caring as a virtue. When emerging technologies are used as supplements for human caring, for instance when a robot brings the medicine could be seen as a caring task being completed, yet the human touch that a patient would need might be missing. In certain contexts of big data technologies, such as healthcare, therefore this technomoral virtue could also be interesting to consider.

Civility: "Making Common Cause"

Vallor explains civility as the "sincere disposition to live well with one's fellow citizens of a globally networked information society: to collectively and wisely deliberate about matters of local, national and global policy [...] and to work cooperatively towards those goods of technosocial life that we seek and expect to share with others." Vallor argues that civility constitutes "self-restrained and polite engagement". This value seems to resonate with others, such as empathy, humility and self-control, yet this value also implies mutually respectful cooperation, which also during the use of privacy-preserving technologies can be seen as an asset.

Flexibility: "Skilful Adaptation to Change"

In Vallor's view technomoral "flexibility is the child of the liberal virtue of tolerance that aims to enable the co-flourishing of diverse human societies." She underlines however that flexibility should be mutually cultivated virtue which seem to pre-require such other values, as mutual respect, empathy, humility and perhaps in cases self-control as well.

Perspective: "Holding on to the Moral Whole"

Cultivating perspective as a virtue, Vallor argues is necessary, because practicing such a value implies that we acknowledge that the emerging technologies we continuously develop have implications that are global in scale and perhaps unforeseeable in depth. Consequences of emerging technologies increasingly magnify and the socio-technical choices we make are accompanied with increasingly complex value-prioritizations. Vallor argues this being unparalleled phenomena in history. Therefore bearing in mind a perspective that our techno-social choices have far-reaching implications is a virtue to embrace.

Magnanimity: "Moral Leadership and Nobility of the spirit "

Vallor describes, that "the magnanimous are those who have rightly deserved the trust of others who can inspire, guide, mentor and lead the rest of us at least towards the vicinity of the good". Magnanimity is therefore a quality of someone who can by example be a moral leader especially during techno-social change.

2.2.2. List of values from Value-Sensitive Design

A second approach to map ethical issues may be found in the so-called 'value-sensitive design' approach concerning technological innovations, including new ICTs.[36] Value-sensitive design regarding communication technologies means that designers should accounts for moral values already during the designing process of technologies. Friedman, Kahn and Borning in their work define a set of "human values with ethical import" which they suggest not as a comprehensive list, but as a set of values which seem to be useful as guiding principles in order to develop value-sensitive ICT for human use. When selecting these values they rely on an extensive list of scholarly literature that advocates for the implementation and active use of one or more of these values. The list of values from value-sensitive design is as follows[37]:

Human Welfare

"refers to people's physical, material, and psychological well-being."

[36] About this see, for instance: Friedman, B., Kahn, P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems. In N. P. Zhang & D. Galletta (Eds.), *Human-Computer Interaction in Management Information Systems: Foundations* (pp. 1–27). New York: M. E. Sharpe. Retrieved from https://cseweb.ucsd.edu/~goguen/courses/271/friedman04.pdf
Manders-Huits, N. L. J. L., & Van den Hoven, J. (2009). The Need for a Value-Sensitive Design of Communication Infrastructures. In P. Sollie & M. Duwell (Eds.), *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments.* Boston: Springer.
Nissenbaum, H. (2010). *Privacy In Context Technology Policy And The Integrity Of Social Life*.
Van den Hoven, J. (2007). ICT and Value Sensitive Design. In V. Goujon, P.; Lavelle, S.; Duquenoy, P.; Kimppa, K.; Laurent (Ed.), *IFIP International Federation for Information Processing, The Information Society: Innovations, Legitimacy, Ethics and Democracy* (Vol. 233, pp. 67–72). Boston: Springer.

[37] This table is from Batya Friedman, Peter H. Kahn, Alan Borning (2006) *Value-sensitive design in Information Systems* in (Zhang, N. P., Galletta, D. eds.) Human-Computer Interaction in Management Information Systems: Foundations, M. E. Sharpe publishing, New York, pp. 17-18.

Grant Agreement number: 731873

Ownership and Property
"refers to a right to possess an object (or information), use it, manage it, derive income from it, and bequeath it."

Privacy
"refers to a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others."

Freedom from Bias
"refers to systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias."

Universal Usability
"refers to making all people successful users of information technology"

Trust
"refers to expectations that exist between people who can experience good will, extend good will toward others, feel vulnerable, and experience betrayal."

Autonomy
"refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals."

Informed Consent
"refers to garnering people's agreement, encompassing criteria of disclosure and comprehension (for "informed") and voluntariness, competence, and agreement (for "consent")."

Accountability
"refers to the properties, that ensures that the actions of a person, people, or institution may be traced uniquely to the person, people, or institution. "

Courtesy
"refers to treating people with politeness and consideration."

Identity
"refers to people's understanding of who they are over time, embracing both continuity and discontinuity over time."

Calmness
"refers to a peaceful and composed psychological state."

Environmental Sustainability
"refers to sustaining ecosystems such that they meet the needs of the present without compromising future generations."

According to Friedman, Kahn and Borning the first nine values can be considered as general requirements to embrace in design environments, whereas courtesy, identity, calmness and environmental sustainability are according them specific value-requirements for technology design in general. For

Grant Agreement number: 731873

privacy-preserving technologies a large set of these values seems also highly useful to encompass both during design but also along the course of application processes.

2.2.3. List of values from 'anticipatory technology ethics'

A third approach towards identifying ethical issues for emerging technologies is with the use of so-called 'anticipatory technology ethics'. Brey developed a theory in relation to a diversity of other technology ethics literature, such as ethical technology assessment, techno-ethical scenarios and the so-called ETICA approach. He offers three stages of technology development: the level of the technology, the artefact, and the application. He distinguishes these levels from the perspective of how values can be integrated: a technology level, artefact level and the application level. He defines 'technology' as being "defined, independently of any artifacts or applications that may result from it"; 'artefact' (including functional systems and procedures) being the things that are "on the basis of a technology… developed"; 'applications' being "the particular ways of using an artifact or procedure, or […] particular ways of configuring it[application] for use". For each of these stages he offers a reference list to uphold from ethical perspectives.

The list of values from anticipatory technology ethics is as follows:[38]

Harms and risks

This value includes different harms and risks that new technologies may contain. This value is closely related to the non-maleficence principle ('do not harm'). Typical harms may include health and bodily harm, pain and suffering, psychological harm, harm to human capabilities, environmental harm and harms to society.

Rights and Freedoms

The freedom of movement, freedom of speech and the freedom of assembly Brey considers being essential in a democratic society and they are also indispensable to cultivate other values including those affected by digitalized interactions. For instance, detrimental implications of digital profiling, such as discrimination, unfair treatment or stigmatization, can put the freedom of movement, freedom of speech and even freedom of assembly under pressure.

Autonomy

In view of Philip Brey, autonomy is the ability to think one's own thoughts and form one's own opinions, and the ability to make one's own choices. In his view autonomy prerequires responsibility and accountability and in the digital era it should also include informed consent. Autonomy is therefore intertwined with the above-mentioned freedoms.

[38] This list is from Philip Brey (2012) Anticipatory ethics for emerging technologies, Nanoethics 6(1): 1-11.

Human dignity

Human dignity includes both self-respect and respect towards all humans by humans without any interests.

Privacy

Under privacy Brey refers to notions of privacy that are known from the data protection perspectives, such as information privacy, but also to such notions as bodily privacy which is the more physical extension of privacy in daily life. Another consideration about privacy is that it is relational.

Property

Property as a value includes here both the right to property of someone as well as rights to intellectual property. When referring to property however Philip Brey also underlines the relevance of other human rights such as the right "to life, to have a fair trial, to vote, to receive an education, to pursue happiness, to seek asylum, to engage in peaceful protest, to practice one's religion, to work for anyone, to have a family, etc." as potentially being under pressure in contexts where emerging technologies appear. Yet, these rights are considered here as part of the anticipatory ethics list, the legal section will deal with such rights.

Animal rights and animal welfare

These rights are especially under pressure when animals are used for testing purposes, such as during the development of new medicines or food sorts or even nanotechnology. Given animals are marginally used for testing purposes when developing big data technologies, we will not consider these rights for big data specifically. Yet, we acknowledge that these rights and animal welfare in general are important to uphold and should be considered when big data technologies have indirectly, detrimental implications on them.

Justice (distributive)

Within anticipatory technology ethics justice involves the just distribution of primary goods, capabilities, risks and hazards, nondiscrimination and equal treatment relative to age, gender, sexual orientation, social class, race, ethnicity, religion, disability, etc. Furthermore, it also involves the geographical dimensions, such as north-south justice as well as age-related aspects, such as intergenerational justice. In general justice in anticipatory technology ethics encompasses social inclusion.

Well-being and the common good

This value is perhaps the most general one of all on this list. Well-being means here being supportive of happiness, health, knowledge, wisdom, virtue, friendship, trust, achievement, desire-fulfillment. This also includes being supportive of vital social institutions and structures, democracy and democratic institutions and of culture and cultural diversity.

Grant Agreement number: 731873

2.3. List of values from biomedical ethics

Beyond the area of philosophy of technology the domain of biomedical ethics is also renowned for its list of ethical issues. Tom Beauchamp and James Childress[39] with their book Principles of Biomedical Ethics from 1979 established far-reaching influence in medical ethics and beyond,[40] mostly because of the broad usefulness of their convincing moral theory for healthcare. In their book they distinguish four main principles: autonomy, non-maleficence, beneficence and justice as the most essential guidelines to consider within patient and medical professional relations. Importantly Beauchamp and Childress perceive these principles in light of prima face duties. In other words, they argue that a continuous 'duty of care' should sustain these principles during any action in healthcare. These principles can be part of prioritizations but they need to be considered as prime convictions for any healthcare professional while carrying out his/her care duties. The list of values from biomedical ethics is as follows:

Autonomy

is considered to be the right of the individual of making his or her own decision or choice.

Beneficence

is described as the principle of acting with the best interest of the other in mind.

Non-maleficence

In their argument non-maleficence is rooted in the Hippocratic Oath, and means 'above all, do no harm' to others.

Justice

is described as the fair and equal treatment of others. They especially stress the importance of a continuous aspiration for fairness and equality during (healthcare) treatments.

These values originally defined for biomedical ethics are also thought-provoking for the context of big data. Although autonomy and justice came up already in earlier-mentioned lists (of values), the novelty of the principles of beneficence and non-maleficence could provide additional benefit during algorithmic decision-making processes by big data. Perhaps upholding the principle of beneficence which would require acting always in the best interest of others is somewhat an ambitious requirement for big data-based practices. Yet, the principle that requires above all not harming other, the non-maleficence principle, could be regarded as a quite useful moral virtue to follow especially during complex big data-driven practices.

---

[39] Beauchamp, T. L., & Childress, J. F. (2012). *Principles of Biomedical Ethics* (Seventh Ed). New York: Oxford University Press.

[40] McCormick, T. M. (2013). Principles of Bioethics. *Ethics in Medicine*. Retrieved from https://depts.washington.edu/bioethx/tools/princpl.html.

Page, K. (2012). The four principles - Can they be measured and do they predict ethical decision-making? *BMC Medical Ethics*, *13*(10).

Waltho, S. (2006). Response to Westin and Nilstun. *Health Care Analysis*, *14*(2).

Westin, L., & Nilstun, T. (2006). Principles help to analyse but often give no solution - secondary prevention after a cardiac event. *Health Care Analysis*, *14(2).*

These four core values still provide the basis for ethical codes and trainings in medicine.[41] Yet, Snyder and Gautier[42] extended this list by two other values: the principle of respect for dignity and the principle of veracity. They argue these principles during medical treatments are of additional value.

Respect for dignity

For instance, the respect for dignity is considered as a value that should be applicable even to patients who are not able to take conscious decisions anymore.

Veracity

The principle of veracity is described when a 'capable patient' needs to acquire an as complete as possible 'truth' knowledge about his or her condition. They argue that only such a complete knowledge can render a patient into a position where he or she can execute a well-informed decision about any possible treatment.

2.4. Ethical value considerations for techno-social change

After having taken a dive into the existing literature on ethical values we distilled from the schemes in the previous section a total of ten values we think are most under pressure within the context of big data technologies, see Table 2. For this, we use the so-called 'ethical matrix', in which three general principles are defined that can be useful to further structure ethical concerns in society.

Mepham calls these the pluralism of principles: *care for well-being; respect for dignity;* and *justice*. In his definition, these three principles can be regarded as overarching guidelines under which other values can be selected. We used these three principles to categorize the values from the four lists of values we introduced above.

[41] Page, K. (2012). The four principles - Can they be measured and do they predict ethical decision-making? *BMC Medical Ethics*, *13*(10).
Price, J., Price, D., Williams, G., & Hoffenberg, R. (1998). Changes in medical student attitudes as they progress through a medical course. *J Med Ethics*, *24*(2), 110.

[42] Synder, J. E., & Gauthier, C. C. (2008). The Underlying Principles of Ethical Patient Care. In *Evidence-based Medical Ethics* (pp. 11–17). Humana Press.

Grant Agreement number: 731873

Mepham's pluralism of principles[43]	*Technomoral virtues*[44]	*Values from value-sensitive design(VSD)*[45]	*Values from Anticipatory technology ethics* [46]	*Values in biomedical ethics*[47]	*e-SIDES: values for big data technologies*
Care for well-being	**Care**	**Human Welfare**	**Well-being and the common good**	**Beneficence**	**Human welfare**
	Magnanimity, Courage	**Autonomy**	**Autonomy**	**Autonomy**	**Autonomy**
	Humility, Self-control	**Calmness**	**Health, (no) bodily and psychological harm**	**Non-maleficence**	**Non-maleficence**
Respect for justice	**Justice**	**Freedom from Bias; Universal usability**	**Justice (distributive)**	**Justice**	**Justice (incl. equality, non-discrimination, digital inclusion)**
	Perspective	**Accountability**	**N/A**	**N/A**	**Accountability (incl. transparency)**
	Honesty, Self-control	**Trust**	**N/A**	**Veracity**	**Trustworthiness (including honesty and underpinning also security)**
Respect for dignity	**N/A**	**Privacy; Informed Consent; Ownership and Property**	**Rights and freedoms, including Property**	**N/A**	**Privacy**
	Identity	**Identity**	**Human dignity**	**Respect for dignity**	**Dignity**
	Empathy, Flexibility, Courage, Civility	**Courtesy**	**N/A**	**N/A**	**Solidarity**
	Courage, Empathy	**Environmental Sustainability**	**(No) environmental harm, Animal welfare**	**N/A**	**Environmental welfare**

Table 2 Virtues to uphold during techno-social change and specifically regarding big data technologies

Derived from the above lists of values (the technomoral virtues, the values indicated by value-sensitive design and anticipatory emerging technology ethics and those defined in biomedical ethics) we distinguish

[43] Mepham, B. (2010) 'The Ethical Matrix as a Tool in Policy Interventions: The Obesity Crisis', in (F-T. Gottwald et al. eds) Food Ethics, Springer Science Business Media, pp. 17-28

[44] Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, pp. 120-121

[45] Friedman, B. et al. (2006) 'Value Sensitive Design and Information Systems' in (Zhang, N. P. and Galletta, D. eds.) Human-Computer Interaction in Management Information Systems: Foundations, M.E. Sharpe Publishing, pp. 19

[46] Brey, P. (2012) Anticipatory Ethics for Emerging Technologies, Nanoethics 6(1), 1-13

[47] Beauchamp, T. and Childress, J. (2012) Principles of Biomedical Ethics, 7th edition, New York, Oxford University Press, pp.

here ten values that we consider highly important to uphold within the contexts in which big data technologies are developed and used.

The division and clustering of values according to Mepham's pluralism of principles: care for well-being, respect for dignity and respect for justice is based on arbitrary choices and the categorization could have been done otherwise. Yet, the present clustering seems helpful to pinpoint what virtues or values can be used best for ethical assessments of big data technologies. The list of values that we distilled from all of the above lists is as follows:

Human welfare

Human welfare is perhaps one of the most straightforward values or virtues to protect.[48] This value, although described in different ways, came to the foreground in all value lists. Value-sensitive design distinguishes it specifically. Anticipatory technology ethics refers to well-being and the common good in general, but we see these values closely related to human welfare. In biomedical ethics when upholding the value of beneficence – doing well to others - human welfare can also be seen as something that will be preserved by it. Furthermore, in the list of technomoral virtues human welfare can rather be seen as related to the value of care in general. In relation to care Vallor points out that some of the biggest dangers of emerging big data technologies for human welfare can appear when "emerging technologies are used as supplements for human caring".[49] This value should therefore also be considered when assessing big data technologies. As Vallor explains care - but we consider human welfare in general - can be under pressure when big data technologies begin to take over healthcare tasks from humans. Human welfare as a value can also be jeopardized when in other contexts big data technologies project discriminative images about humans to other humans, such as the contexts of employment, schooling or travelling exemplify instances of unfair treatment and discrimination of citizens.

Autonomy

The word autonomy finds its roots in Greek philosophy and stems from the words 'auto-nomos' referring to city states which were 'self-governing'. Only during the period of the European Enlightenment[50] became autonomy used as a personal property.[51]

Although autonomy has different forms in philosophical thinking, such as moral autonomy, existentialist autonomy, relational autonomy, personal autonomy or autonomy seen as a right. Since our focus lies on the implications of big data technologies, the focus of this chapter is on personal autonomy and autonomy seen as a right. Yet, the concept of relational autonomy shall be shortly described here as it relates to virtue ethics.

---

[48] Hurthouse, R. (2006) Applying Virtue Ethics to Our Treatment of the Other Animals in Welchman, J (eds.) The Practice of Virtue, Hackett Publishing Company.

[49] Vallor S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, pp. 119.

[50] Rawls, J. (1971). A Theory of Justice, Revised edition (1999) Cambridge, MA: Harvard University Press. Christman, John and Joel Anderson, eds. (2005). Autonomy and the Challenges to Liberalism: New Essays, New York: Cambridge University Press.

[51] See footnote 29.

Grant Agreement number: 731873

Relational autonomy is a spinoff of feminist critique of traditional autonomy.[52] This conception replaces individualistic views of autonomy by conceptions on the self and deliberation and reasoning. This refers to the form of autonomy which means self-governance and where the self is constituted by relationships to others. Since big data creates a dense network between systems and humans we consider this aspect of autonomy crucial when analysing contexts of big data technologies.

Autonomy as a value is also prominent on the value lists of VSD, anticipatory emerging technology ethics and biomedical ethics. Vallor's technomoral virtue of magnanimity and courage – daring to take risks - seem closely related to autonomy. Magnanimity or moral leadership, for instance, in Vallor's view is a virtue that everyone needs to cultivate in him/herself, but especially those who shape trends for emerging big data technologies: technology designer, entrepreneurs and others. An autonomous person in our big data society, therefore, desirably needs to be a magnanimous and courageous one in his/her activities. In this fashion, such an autonomy being a broad value in its scope would also serve such other "pluralism of principles"[53] as (both biological and technology-mediated) well-being.
Autonomy, however, is at risk when big data technologies and data transfer processes are not transparent for citizens, when decision-making takes place beyond notifying them about how a decision came about. Furthermore, autonomy can also be in danger when big data-driven profiling practices limit free will, free choice and turn out to be rather manipulative instead of supportive in raising awareness or cultivating knowledge about, for instance, news, culture, politics and consumption.

Non-maleficence

Non-maleficence is a core value to uphold in biomedical ethics.[54] However, we have selected it as an essential one also when considering big data technologies. It might not seem straightforward, but detrimental implications of big data-based calculations could be highly limited if non-maleficence as a design requirement would be implemented into big data processes. Humility and self-control from technomoral virtue ethics lists are intertwined with non-maleficence, as much as calmness from the value-sensitive design list, and health (protection from both bodily and psychological harms) from the anticipatory emerging technology list. With respect to humility, for instance, when using big data Vallor suggests that we put down our ego and accept that we will not have a continuous overview upon how data within algorithmic networks flow and what their consequences will be. This also means that in terms of big data technologies non-maleficence could be regarded as a value under continuous pressure. For instance, possibilities for data reuse in the world of big data are vast and certain limits are often only set by law. For many situations and developments there is no defined legal framework (yet), even though data reuse would technically be possible for such purposes as earning profit. Hence, non-maleficence, self-control and humility could be regarded as highly important virtues in jeopardy.

---

[52] Friedman, M. (1998). "Feminism, Autonomy, and Emotion," in Norms and Values: Essays on the Work of Virginia Held, Joram Graf Haber, ed., Lanham, MD: Rowman and Littlefield;
Baumann, Holgar (2008). "Reconsidering Relational Autonomy. Personal Autonomy for Socially Embedded and Temporally Extended Selves," Analyse and Kritik, 30: 445–468.
[53] See footnote 21.
[54] Beauchamp, T. L., & Childress, J. F. (2012). *Principles of Biomedical Ethics* (Seventh Ed). New York: Oxford University Press.

Grant Agreement number: 731873

Justice

Whereas justice as a value can be considered as a prime requirement in everyday life, this value appears almost on all lists of values. Value-sensitive design, however, refers rather to freedom from bias, but we consider this highly overlapping. Friedman and her co-authors underline the essential need to prevent or remedy "systematic unfairness" as a form of bias. Within the context of big data, systematic unfairness emerges, for instance, by generating false positives during preventative law enforcement practices or false negatives during biometric identification processes. Such instances put constant pressure on the value of justice. Given that big data accelerates the speed and quantity of data transfer and creates immense possibilities for data reuse, instances for false identity verification, false accusation or stigmatization, for instance as a consequence of big data-led correlations can also increase.

Accountability

Accountability is a virtue that requires constant assessment in democratic societies.[55] From all the four lists of virtues resulting from our desk research, accountability is explicitly stated only on the list of value-sensitive design. The technomoral virtue 'perspective' seems to encompass essential qualities for upholding accountability. As Vallor argues 'perspective' means the acknowledgement that emerging technologies will have implications that are "global in scale and unforeseeable in depth". Such a stance and awareness can be regarded as getting us half way to develop accountability checks in big data-based interactions. This is so, because they point out that transparency is a prerequisite of accountability and it is under constant pressure in domain of big data. For instance, in the healthcare domain patients or in the marketing domain consumers often do not know what it means when their data is shared via surveys for research and marketing purposes. We consider therefore that accountability is a highly relevant value that can be under pressure in many contexts of big data.

Trustworthiness

Trustworthiness and honesty are virtues that requires mutual caring especially in a networked world.[56] The list of values in VSD explicitly contains trust as a value, whereas the list of technomoral virtues refers rather to honesty, anticipatory technology ethics does not mention trust or honesty. Biomedical ethics lists veracity as a value, which we consider highly related to trustworthiness.

Given the context of big data, especially the instances when algorithms are used for manipulation, honesty as a value to embrace seems highly relevant. However, the parameters of one general truth are quite difficult if not impossible to outline as Vallor points out, because the truth content of a message can change by to whom we speak, where we gather our information or how we present it. Still the intention of speaking with honesty within the context of big data should remain practiced. When talking about trustworthiness and honesty the value of veracity in biomedical ethics seems also highly related. The principle of veracity when implemented as a moral requirement for big data collectors, data brokers, data

---

[55] Braithwaite, John. (2006) "'Accountability and Responsibility through Restorative Justice'". In Public Accountability: Designs, Dilemmas and Experiences, Edited by: Dowdle, M. 33–51. Cambridge: Cambridge University Press.

Jos, Philip H. and Tompkins, Mark E. (2004) 'The Accountability Paradox in an Age of Reinvention'. Administration and Society, 36(3): 255–81.

[56] Keymolen, E.L.O. (2016). Trust on the line: a philosophycal exploration of trust in the networked era. Erasmus University Rotterdam.

Grant Agreement number: 731873

scientists and other stakeholders could, for instance, also improve the legal position of the individual citizen whose data provides resources for big data-driven calculations. Prescribing the practice of veracity for stakeholders during big data-based processes could, for instance, strengthen the enforcement of the aim to have informed consent; the right not to be subject to profiling or the right to explanation for individual citizens as these will be new EU data protection requirements to adhere to. The latter rights seem to be constantly under pressure when veracity, honesty and trust in general are not upheld, for instance, because big data transfer processes are not transparent, citizens do not know where their data are and where to turn to for remedy especially when big data-based calculations are made about them.

Privacy

Only value-sensitive design lists privacy as a separate value to cherish. Anticipatory emerging technology ethics includes among other rights and freedoms the fundamental right to privacy. The value lists of biomedical ethics and of the technomoral virtues do not refer to privacy explicitly, but privacy can be related to the above-mentioned value of dignity and identity in many ways. Given the context of big data we consider this value as essential to uphold. Since the rise of the internet there is extensive literature on privacy.[57] Privacy as a value we consider here being broader than its notion often referred to in the legal, data protection domain (see Chapter 3). Privacy as a fundamental value is the closest in its broad scope to privacy as a fundamental right. Privacy as fundamental value first of all includes respect for others, and specifically respect for someone's private sphere, private conversations, writing – e.g. confidentiality of mail – and also any actions that one keeps intentionally unexposed to the broad public. A cornerstone of privacy as a virtue lies in respecting the boundaries someone else has drawn him/herself and the boundaries one would also like to see being protected from intrusion by others. These instances of privacy within big data-mediated interactions are in constant jeopardy, as persons can be identified in datasets increasingly easily even if technical measures of anonymization are in place. Simply the myriad of correlations between personal data in big data schemes allows for easy identifiability. Such instances of big data technologies, therefore, provide many ways to disrespect one's private interactions and relations.

[57] Moerel L, Prins J.E.J., Hildebrandt, M., Tjong Tjin Tai, T. F. E., Zwenne, G. J. en Schmidt, A. H. J. (2016) Homo Digitalis, Wolters Kluwer Publishing
Hildebradt, M. de Vries, K. (2013) Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the philosophy of technology, Routledge
De Hert P., Gutwirth, S. (2006) 'Privacy, data protection and law enforcement. Opacity of the individual and transparency of power' in E. Claes, A. Duff & Gutwirth, S. (eds..), Privacy and the criminal law, Antwerp/Oxford, Intersentia, 61-104.
Custers B.H.M. & Ursic H. (2016), Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection, *International Data Privacy Law* 6(1): 4-15.
Omer, T. and Polonetsky, J. "Privacy in the Age of Big Data: A Time for Big Decisions." February 2, 2012. 64 Stan. L. Rev. Online 63. http://www.stanfordlawreview.org/online/privacy-paradox/big-data (last visited June 28, 2012)
Koops, B-J. and Newell, B.C. and Timan, T. and Škorvánek, I. and Chokrevski, T. and Galič, M, A Typology of Privacy (March 24, 2016). University of Pennsylvania Journal of International Law 38(2): 483-575 (2017); Tilburg Law School Research Paper No. 09/2016. Available at SSRN: https://ssrn.com/abstract=2754043.

Dignity

Dignity[58] as a value can be found on the list of values in anticipatory emerging technology ethics and biomedical ethics. On the lists of technomoral virtues and value-sensitive design, identity seems the closest in notion and there is some overlap here. When dignity is preserved, identity is also preserved and the same can be said about dignity when identity is preserved. Human dignity can therefore be regarded as a prime principle since all implications of big data technologies on humans affect this value in one way or another. For instance, both privacy and identity concern dignity and can be regarded as relational concepts. If they are under pressure, they also affect other values.

Dignity appears fragile when big data technologies are in place. For instance, when revealing too much about a user, principles of data minimization and design requirements of encryption appear to be insufficient. Adverse consequences of algorithmic profiling, such as discrimination or stigmatization, also demonstrate this. Implementing the principle of respect for dignity into design processes of big data technologies could minimize such diverse effects.

Solidarity

Solidarity as a value does not appear on any of the value lists referred to above. Yet, we consider solidarity as a highly relevant value in domains of big data. Technomoral virtues such as empathy, flexibility, courage and civility all relate to solidarity as these values can be seen as building blocks of an attitude of solidarity. Solidarity, however, is a value which has been embraced to different degrees over time. For instance, before the fall of the iron curtain in Poland, solidarity became a value that assembled crowds, creating a new positive ideology of compassion and brotherhood against decades of Soviet political oppression. This symbolic but also enacted form of solidarity in today's circumstances, at least in Western Europe, cannot be comparable as there is no oppression. Yet, the implications of showing/expressing solidarity in daily life by having the courage to come up for others and having the flexibility – the latter being closely related to the liberal virtue of tolerance according to Vallor – in order to "enable the co-flourishing of diverse human societies" is crucial to uphold during discussions on big data. Many instances of big data-based calculations in which commercial interests are prioritized rather than non-profit-led interests, are examples of situations in which solidarity is under pressure. When, for instance, immigrants are screened by big data-based technologies, they may not have the legal position to defend themselves from potential false accusations resulting from digital profiling. These are examples of solidarity being under pressure during big data-based interactions.

Environmental welfare

This value appears as environmental sustainability on the list of value-sensitive design, as no environmental harm (including animal welfare) on the list of anticipatory emerging technology ethics and is not listed in biomedical ethics. Among the technomoral virtues, empathy and courage seem to be the most closely related to it. Although empathy has not been developed as a concept concerning non-

[58] Tischner, J. (2005) The Ethics of Solidarity, Retrieved on 22nd of August from < http://www.tischner.org.pl/Content/Images/tischner_3_ethics.pdf>
Düwell, M. (2017) Human Dignity and the Ethics and Regulation of Technology. Brownsword, R., Scotford, E., Yeung K. (Eds.) The Oxford Handbook of Law, Regulation and Technology, Oxford Handbooks Online.

Grant Agreement number: 731873

humans, the environment (including animals), for instance, in Japanese[59] and tribal philosophies, such as the Native American culture[60] or the South-African Ubuntu[61] culture demonstrates is highly respected, potentially as much as human life. Although big data has rather indirect effects on the environment, the current rush for lithium in Latin-America[62], which is the critical ingredient of all batteries on the world, shows how environmental welfare as a value is under pressure by big data technologies.

Certainly our list including these ten values, summarized in Table 1, is not exhaustive, but these ten value dimensions provide, relatively broad, yet relevant perspectives for discussions concerning the implications what big data technologies bring along for our current societies. Human welfare, autonomy, non-maleficence, justice, accountability, trust, privacy, dignity, solidarity, environmental welfare are all values that are constantly under pressure within the context of big data. Hence, this selection seems useful for us to provide ethical stepping stones to assess privacy-preserving big data technologies in the next deliverables.

[59] Callicott, B. J. & McRae, J. (Eds.) (2017) Japanese Environmental Philosophy, Oxford University Press.
[60] Booth, A. L. (2008) Environment and Nature: The Natural Environment in Native American Thought in Selin H. (ed.) 'Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures' pp. 809-810, Springer The Netherlands
[61] Chuwa L. (2014) Ubuntu Ethics. In: African Indigenous Ethics in Global Bioethics. Advancing Global Bioethics, vol 1. Springer, Dordrecht.
[62] Frankel, T. C. & Whoriskey, P. (2016) Tossed aside in the 'white gold' rush Indigenous people are left poor as tech world takes lithium from under their feet, The Washington Post <http://www.washingtonpost.com/graphics/business/batteries/tossed-aside-in-the-lithium-rush/>

Grant Agreement number: 731873

3. Legal perspective

Relevant human rights	ECHR	EU Charter	List of legal issues
Right to Privacy	Art. 8	Art. 7	Lack of transparency Vagueness of the concept of harm and lack of individually attributable harm Proportionality Accountability Establishing the adequate regulatory framework The role of private actors in the context of human rights framework
Right to personal data protection	N/A	Art. 8	
Freedom of expression	Art. 10	Art. 11	
Freedom of assembly and association	Art. 11	Art. 12	
Right to non-discrimination	Art. 14	Art. 20 Art. 21	
Right to effective remedy and fair trial	Art. 6 Art. 13	Art. 47 Art. 48	
Consumer protection	N/A	Art. 38	

*Table 3 O*verview of the identified legal issues

Grant Agreement number: 731873

3.1. Introduction

Ground-breaking technological developments do not leave the legal realm unconcerned; they necessitate a certain legal approach – identification of the loopholes, inadequacies, but also opportunities for intervention and shaping the technological realm. Big data technologies are undoubtedly such a ground-breaking phenomenon.[63] Bearing many promises for scientific progress, they set to impact the individuals and also societies as a whole. They might facilitate many processes of different natures but their real impact could be also truly transformative, possibly also in fundamentally adverse manner. At the nexus of key ethical values (see Chapter 2) and the big data 'revolution', fundamental rights provide for the legal framework with great potential in addressing such new technological challenges.

Drawing on the findings of the previous chapter, which sets forth the list of ethical values relevant in the context of big data technologies, this chapter focuses on the legal emanations of such key ethical values, taking the European framework of fundamental rights as a point of reference. Thus, the aim of this section is to analyse the legal framework for the protection of fundamental rights from the perspective of the distinctive challenges brought about by the developments of the big data technologies and their applications.

The chapter looks at the legal issues connected to the development of big data technologies in the realm of fundamental rights. The legal issues concern: (1) the norms set forth in the discussed human rights instruments as well as the relevant case law concerning the substantive scope of protection of selected rights, as applied in the context of big data technologies and (2) the functioning of the regulatory framework in general in this context. Outside the scope of this chapter are the issues connected to discussion of the moral values underpinning the rights as well as issues focused purely on the enforcement of discussed rights.

This chapter starts with delineating the main characteristics of the relevant European legal framework. This legal framework consists of the European Convention of Human Rights (ECHR)[64] and the EU Charter of fundamental rights (the EU Charter).[65] These legal frameworks for the protection of fundamental rights are described in Section 3.2 in order to clarify the scope of application of each of these instruments, to explain judicial interpretation techniques and to describe the interrelation between different legislative measures. Next, Section 3.3 proceeds to analyse the key distinctive characteristics of the big data technologies with reference to the case law of both the European Court of Human Rights (ECtHR) and the Court of Justice of the European Union (CJEU). The distinctive features of big data technologies are introduced along two axes: the particularities of the processing technology itself (focusing on the scale, velocity and variety of sources for data processing) as well as distinct stages of the big data technologies applied in practice (i.e., gathering of the data, processing, data based decision-making). Section 3.3 also describes the catalogue of the fundamental rights provided by the ECHR and the EU Charter which are at

[63] See for example: Akerkar, R., G. Lovoll, S. Grumbach, A. Faravelon, R. Finn and K. Wadhwa (2015) 'Understanding and mapping Big Data', deliverable 1.1 byte project.
[64] European Convention for the Protection of Human Rights and Fundamental Freedoms, 1953.
[65] European Union, Charter of Fundamental Rights of the European Union [2000] OJ C364/01 and [2010] OJ C83/389.

stake in the context of big data technologies and applications and subsequently scrutinizes how and at which stages big data technologies could potentially interfere with these rights.

3.2. European fundamental rights framework

3.2.1. Human rights as legal norms

There is extensive literature[66] in the field of political philosophy on what differentiates human rights from other moral and legal rules, what is their nature, content, grounds for existence and legitimacy. The purpose of this section is to look at human rights from the perspective of their functioning within a legal framework – as legal norms. In this context the system of human rights can be perceived as 'a powerful instrument for realizing moral values'.[67] According to Buchanan, 'human rights law, like law generally, is an institutionalized form of practical reasoning that serves moral values'.[68] The institutionalisation of human rights law in the form of different legal instruments provides for a framework of reference for the realisation of moral values by the state, making these values enforceable as legal rights. Thus, the instruments incorporating human rights allow the right holders (the individuals) to enforce certain moral values enshrined in the human rights against the duty bearers (the states).

While human rights as legal norms impose certain obligations on the duty bearers, they should not be treated as rules to be unequivocally obeyed, but rather as what Alexy calls 'optimization requirements', i.e., norms that should be realized to greatest extent all things considered.[69] Thus, most human rights are not absolute rights but are applicable pursuant to the principle of proportionality to a certain, maximally available in given circumstances, degree.

The focus of this chapter is in particular on the human rights framework within Europe, or more precisely, within the European Union, where two human rights instruments are of particular relevance: the European Convention of Human Rights (ECHR)[70] and the Charter of Fundamental Rights of the European Union (EU Charter).[71] The following subsections discuss briefly the main characteristics of these legal instruments.

[66] For example: C. Corradetti, *Philosophical Dimensions of Human Rights*, New York: Springer, 2012; R. Cruft, *Human Rights as Rights*, in G. Ernst and J. Heilinger, (eds.), The Philosophy of Human Rights: Contemporary Controversies, Berlin: Walter de Gruyter, 2011.
[67] A. Buchanan, *The Heart of Human Rights*, OUP Online Scholarship 2014, p. 5.
[68] *Ibidem*
[69] R. Alexy, *A Theory of Constitutional Rights*, 2002,p. 47.
[70] European Convention for the Protection of Human Rights and Fundamental Freedoms, 1953.
[71] European Union, Charter of Fundamental Rights of the European Union [2000] OJ C364/01 and [2010] OJ C83/389.

Grant Agreement number: 731873

3.2.2. The ECHR framework

Scope of application

The ECHR is an international convention dating from 1953, whose signatories are 47 member states of the Council of Europe. These include all EU member states and several other European countries. Ratification of the convention means that the state is obliged to conform to its provisions in its domestic legal system. This does not mean that the member states must incorporate the provisions of the ECHR into their legal systems, but rather that the substance of the rights and freedoms set forth must be secured under the domestic legal order, in one form or another, to everyone within the jurisdiction of the member states.[72] In case any given member state fails to protect the rights enshrined in the ECHR, an individual whose right was violated as a result can bring the case before the European Court of Human Rights (ECtHR) provided she exhausted all the national instances to seek a remedy.

The decisions of the ECtHR, while being only declaratory, are binding on the member state against which the claim was brought. This means that each member state which was found to have violated the ECHR is obliged to implement the measures aimed at remedying the situation in the immediate case disputed before the ECtHR and to address any systemic deficiencies in the national law that led to infringement of the ECHR.[73] In this context, it has been underlined in the literature that the national courts play a particularly important role in ensuring the application of the national legal framework in accordance with the provisions of the ECHR as interpreted in the ECtHR case law.[74]

It is worth noting that in practice the compliance with the judgments of the ECtHR is very high, meaning that the member states found to infringe any provision of the ECHR generally make necessary adjustments in their national legal framework in order to prevent any further violations of the ECHR.[75]

Positive obligations of a state and indirect horizontal effect

As noted in the previous subsection, the duty bearers in the human rights framework are states. Consequently, in the disputes concerning the violation of a particular right under the ECHR, the defendants are states in which such a violation is claimed. However, it is not always the case that the state infringed the ECHR through its abusive action: sometimes the infringement is a result of member state's failure to protect an individual against the abuses. Thus, the states may not only have a duty to refrain from actions breaching a convention, but also have an active duty to set the legal framework that effectively protects an individual against such breaches by others, i.e., a duty to act. This duty stems from

[72] *Soering v. the United Kingdom*, Application No. 14038/88, para. 120.
[73] A. Caligiuri, N. Napoletano, *The Application of the ECHR in the Domestic Systems,* The Italian Yearbook of International Law Online, Vol. 20, Issue 1, pp. 125 – 159, 2010.
[74] *Ibidem.*
[75] G. Letsas, *A Theory of Interpretation of the European Convention on Human Rights*, OUP Online Scholarship, 2009.

the principle of effectiveness which requires the states to assure that the human rights protected under the ECHR are also actually protected in practice and not only in abstract terms.[76]

Some disputes before the ECtHR originate from cases in which on a national level only non-governmental litigants are involved. A typical example is a case in which a particular press publication infringed the right to privacy of an individual and the claim of such infringement was not recognized in the national proceedings.[77] As noticed by the ECtHR in the context of the right to privacy, although the object of Article 8 of the ECHR (which sees to the right to privacy) is essentially that of protecting the individual against arbitrary interference by the public authorities, it does not merely compel the state to abstain from such interference: in addition to this primarily negative undertaking, there may be positive obligations inherent in an effective respect for private or family life. These obligations may involve the adoption of measures designed to secure respect for private life even in the sphere of the relations of individuals between themselves. Therefore, the human rights under the ECHR might have what is called an *indirect horizontal effect*, meaning that they are imposing on states a positive obligation to provide measures to safeguard human rights in horizontal relations between individuals.

ECHR as a living instrument

Another aspect of the ECHR framework for the protection of human rights which is worth exploring is the interpretative methodology of the ECtHR connected to the idea of the ECHR is a 'living instrument'. Pursuant to the ECtHR case law, this means that the ECHR must be interpreted in the light of present-day conditions,[78] which in turn obliges the ECtHR to be influenced by the developments and commonly accepted standards[79] among the member states or so-called other "signs of evolution of attitudes amongst modern societies".[80] The concept of the ECHR as a 'living instrument' reflects how the ECtHR is impacted by the social trends in assuring the proper level of protection of human rights protected under the ECHR. However, as noted by Letsas in his analysis of the ECtHR case law in this context, it shows that the ECtHR was primarily interested in evolution towards the moral truth of the ECHR rights, not in evolution towards more commonly accepted standard, regardless of its content.[81] Consequently, such an approach of the ECtHR suggests that it undertakes the moral reading of human rights, attributing them certain objective value and the evolving interpretation of the ECtHR allows it to properly protect such objective values regarding the changing circumstances.

However, not every social change would be followed by the ECtHR: this implies that the evolution of the attitudes in modern societies could be embraced or countered by the ECtHR, depending on whether such

[76] See: A. R. Mowbray, The Development of Positive Obligations under the European Convention on Human Rights by the European Court of Human Rights, 2004, p. 221.
[77] *Von Hannover v. Germany*, (application no. 59320/00), para. 45.
[78] *Tyrer v. United Kingdom* (application no. 5856/72), para. 31.
[79] *Ibidem*.
[80] *Ibidem*.
[81] G. Letsas, *A Theory of Interpretation of the European Convention on Human Rights*, Oxford University Press Online Scholarship, 2007, Chapter 3, p. 17.

Grant Agreement number: 731873

evolution goes hand in hand with what the ECtHR considers to be the pre-existing, objective value of the right enshrined in the ECHR.

3.2.3. The EU Charter framework

Sources of fundamental rights in the EU law

As noted by Allan Rosas, judge of the Court of Justice of the EU (CJEU), the development of the human rights framework within the EU was a 'a story of judge-made law'.[82] Indeed, long before any reference to the fundamental rights was included in the treaties that form the basis of EU law,[83] the CJEU recognized that they form part of the general principles of European Community law.[84]

Currently, within the legal framework of the EU several different instruments providing protection of human rights can be traced. Following entry into force of the Lisbon Treaty,[85] in 2009 the European Charter of Fundamental Rights (the EU Charter) became a binding legal instrument with its legal status equivalent to that of the treaties that form the basis of EU law.

Furthermore, certain fundamental rights are explicitly regulated in the secondary legal instruments (typically EU regulations and EU directives), stipulating more detailed legal provisions for the protection of these fundamental rights. For example, the right to data protection (in protected under Article 7 of the EU Charter) has been regulated in more detail in the 1995 EU Data Protection Directive (DPD)[86] which will be substituted in May 2018 by the General Data Protection Regulation (GDPR).[87] Similarly, the right to non-discrimination, the right to access to documents, and the right to consumer protection are among the rights that are further elaborated in the EU secondary law. These rights remain fragmented and limited to fields where the EU is granted a specific competence, as there is no mandate for the EU by its member states to develop a general human right policy.[88]

---

[82] A. Rosas, When is the EU Charter of Fundamental Rights applicable at national level?, *Jurisprudencja*, 2012, 19(4), p. 1270.

[83] The two treaties that form the basis of EU law are the Treaty on the European Union (TEU) and the Treaty on the Functioning of the European Union (TFEU).

[84] Case 29/69 *Stauder.*

[85] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community (OJ C 306, 17.12.2007), entry into force on 1 December 2009.

[86] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, [1995] OJ L 281/31.

[87] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), [2016] OJ L 119/1.

[88] E. Muir, *The Fundamental Rights Implications of EU Legislation: Some Constitutional Challenges*, CMLR 51, 2014.

Scope of application of the Charter

Pursuant to Article 51(1) of the EU Charter, its provisions are binding only to the institutions and bodies of the EU and the member states when they are implementing EU law. Similarly the EU Charter does not extend in any way the competences transferred by the EU member states to the EU. In practice this means that the legal framework provided for by the EU Charter is triggered wherever there is a nexus between a particular human right and some other provision of EU law.[89] Thus, the scope of application of the EU human rights framework is determined by the scope of application of the EU law.[90]

Admittedly, in the recent case law of the CJEU there is a precedent where the scope of the application of the EU Charter seems to be applied broadly. In its landmark ruling in *Fransson*,[91] the CJEU decided that the scope of the EU Charter extended to the situation where the infringing national measure while falling within the scope of the EU law was not an instance of the implementation of the EU law into the national legal system. The CJEU indicated that '[t]he applicability of European Union law entails applicability of the fundamental rights guaranteed by the Charter',[92] meaning that the scope of application of the Charter squarely falls within the scope of application of the EU law.

Similarly, the CJEU further elucidated the wide scope of application of the EU Charter in *Melloni*,[93] where it is indicated that member states can provide a higher level of protection for fundamental rights than that provided for under the EU law provisions (pursuant to Article 53 of EU Charter), but only in cases where a field is not completely regulated under the EU law.

As regards the interpretation of the provisions of the Charter, Article 52(3) stipulates that whenever a right set forth in the EU Charter corresponds to those contained in the ECHR, the scope and meaning of the Charter rights should be the same as that of the ECHR rights. However, the ECHR serves as a minimal threshold of protection and thus the EU Charter rights can afford more extensive protection in that regard.

Horizontal effect of the EU fundamental rights

Several avenues could be pursued in order to claim the applicability of the EU provisions for protection of fundamental rights directly between the individuals. At the outset, it should be noted that the case can be made for the direct application of the secondary legislation in the horizontal relations. Indeed, in its case law within the framework of antidiscrimination law, the CJEU established that certain directives might have direct horizontal effect when giving specific expression to general principles of EU law.[94] By analogy, it could be argued that the Data Protection Directive, giving specific expression to the general principle of data protection, enshrined in Article 7 of the EU Charter, would qualify for the direct

[89] A. Rosas, When is the EU Charter of Fundamental Rights applicable at national level?, Jurisprudencja, 2012, 19(4), p. 1282.
[90] F. Ferraro and J. Carmona, European Parliamentary Research Service, *Fundamental Rights in the European Union The role of the Charter after the Lisbon Treaty*, March 2015.
[91] Case C-617/10, *Åkerberg Fransson.*
[92] *Ibidem*, para. 21.
[93] Case Case C-399/11, *Melloni*.
[94] Case C-144/04, *Mangold v. Helm*; Case C-555/07, *Kücükdeveci v. Swedex*.

application in accordance to this line of CJEU case law.[95] Moreover, upon entry into force in May 2018, the GDPR will become directly applicable.

Moreover, the possible application of the EU Charter to the private law realm could be taken into account. As often clarified in scholarship, in principle the EU Charter, pursuant to Article 6(3) of the Treaty on the EU has the status equivalent to that of the primary law and, hence, could in theory be applied directly. However, pursuant to the limitation of the scope of application of the EU Charter provided for in Article 51(1), its *rationae personae* is limited to the institutions of the EU and member states when implementing the EU law. Such strict textual reading would thus bar further exploration of the issue of EU Charter's direct horizontal applicability. Nevertheless, following the CJEU's willingness to abandon the culprits of purely textual interpretation of the EU fundamental rights legislation, the case for the direct application of the EU Charter has been explored in the following scholarship.

First, as argued by D. Leczykiewicz,[96] the strict formal criteria of direct and indirect horizontal effect, fragmentally and inconsistently applied by the CJEU in the context of fundamental rights, could be replaced by the substantial considerations related to the power relations and information asymmetries of the contractual parties. Under this take on horizontality, the application of the EU Charter would be warranted in the private law realm where one party to the contract exercises considerably greater power, thus shaping and imposing the terms of the contractual relationship unilaterally. Such principled approach to horizontal application of the EU Charter would aim mainly at achieving the goals of social justice and preclude constitutional overreach into private law where contractual freedom of the parties was duly exercised.

In the same vein, E. Frentziou[97] proposed that the issue of the horizontality of the EU Charter could be approached taking into consideration the common goods involved in the protection of fundamental rights and thus understanding such rights not purely as emanations of individualistic claims but also as expression of socially embedded interests. In this perspective, it is important to take into account social impact of the horizontal effect of the EU Charter and verify whether its application could remedy the injustices inherent in developing reality or privatizing public sphere by powerful private actors. Given the fact that private bodies overtake some of the functions commonly associated with the public bodies, and with very little public scrutiny, extending the framework of the EU Charter to these actors could be considered in accordance with the principle of effective protection of human rights.

The approaches to the issue of horizontality of the EU Charter discussed above correspond very well with the challenges of the rise of private power in the context of data processing online, as will be discussed below.

---

[95] As noted by E. Muir: "The possibility of making use of directives giving specific expression to a fundamental right to enhance the effects of EU law will certainly be explored by applicants in the years to come. An interesting test-case could relate to EU data protection law." (although such possibility is not likely to be tested in the view of the expected entry into force of the Regulation), E. Muir, *The Fundamental Rights Implications of EU Legislation: Some Constitutional Challenges*, CMLR 51, 2014.

[96] D. Leczykiewicz, *Horizontal Application of the Charter of Fundamental Rights*, European Law Review 38, 2013.

[97] E. Frantziou, *The Horizontal Effect of the Charter of Fundamental Rights of the EU: Rediscovering the Reasons for Horizontality*, European Law Journal, Vol. 21, No. 5, September 2015, pp. 657–679.

3.3. Catalogue of fundamental rights relevant in the context of big data applications

Contrary to what might be indicated by the phrase 'big data', the implications of such technologies are not only limited to the quantitative scope of data processing but actually make some qualitative changes in the way that processed data can impact different spheres of life of individuals and society as a whole. In this sense, the big data technology has some transformative effects (which will be discussed below) which might trigger and challenge the various aspects of the human rights framework.

The following subsection (subsection 3.3.1) focuses on the distinctive characteristics of big data that are relevant in the context of the EU human rights frameworks of the ECHR and the EU Charter and on different stages of applying big data technologies which can implicate different human rights discussed. The other subsections distil a list of fundamental rights in both the ECHR and the EU Charter that are implicated (subsection 3.3.2) and provide from that a list of legal issues of big data technologies (subsection 3.3.3).

3.3.1. Three stages of applying big data technologies

Big data and related big data technology are not a unitary phenomenon but rather a process aimed at deriving knowledge from processing very large datasets in order to obtain certain knowledge, for instance, information on groups of society in general or profiles of groups or individuals.[98] In particular, three stages in the application of big data technologies can be differentiated:

a) Data collection

b) Data processing (or data mining)

c) Application of the derived knowledge

All these stages may implicate different agents, consist of different actions and might challenge different aspects of fundamental rights. First, the stage of data collection means acquiring massive data sets (related to volume of data) coming from different sources (related to variety), often merging different data sets together, pursuant to the principle the more data the better (or data maximization). The data that these datasets are comprised of may have the characteristics of personal data, but may be also completely anonymous, hence not relating to identified or identifiable natural persons.[99] Such data might be voluntarily contributed by an individual (for example, through use of social media platforms) or gathered without her knowledge (for example, through gathering anonymous data on traffic patterns).

The second stage is connected to algorithmic analysis of the data sets. Such analysis is data driven, meaning that the departing point of the analysis does not depend on prior formulating of research hypotheses. Nor does such processing of data require the discovery of causal relations within different patterns. The underlying assumption is that given the enormous amount of data processed, finding certain

[98] See: Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, adopted on 2 April 2013.
[99] The concept of personal data will be discussed in detail below.

correlation between a particular data and a proxy can already lead to valuable knowledge.[100] For example, purchasing equipment of a given brand might be used as an indicator of propensity to pay higher prices for airline tickets.

In the third stage, the results of the data mining techniques or other types of big data analytics are interpreted and used as a basis for making decisions. In this context it is important to mention that one characteristic of this process is decompartmentalization of different domains, which means that data obtained in one realm can lead to decisions being taken in a completely different realm.[101] Moreover, as a result of combining many different datasets, information that may seem not to reveal any personal details about an individual can after aggregation lead to the discovery of a new knowledge revealing even details of one's personal life that could be qualified as sensitive data. Such practice might lead for example to direct or indirect discrimination, where based on a seemingly neutral proxy (like zip code) individuals are grouped into different categories which correspond with the prohibited grounds for discrimination (like sex or race) and treated dissimilarly to other groups (for example, by being obliged to pay higher insurance rates).

Different stages of application of big data technologies may implicate different aspects of human rights as set forth in the ECHR and the EU Charter. For example, the stage of data accumulation in an obvious way might infringe upon certain aspects of the right to data protection, but may be completely irrelevant from the point of view of the right to non-discrimination. Thus, the next subsection will take into account these three different stages of applying big data technologies.

3.3.2. List of fundamental rights implicated

Distinctive characteristics of big data technologies put certain aspects of the current framework for the protection of human rights in the EU under pressure. Thus, there is a risk that some values underpinning the human rights set forth in legal instruments might be left without protection due to inadequacy of the current legal framework to tackle such new situations where big data technologies are used. This section lists some of the human rights contained in the European Convention of Human Rights (ECHR) and the EU Charter of Fundamental Rights (the EU Charter), explaining in particular the substantive scope of each right as developed in the case law of the European Court of Human Rights (ECtHR) and Court of Justice of the EU (CJEU) respectively, and describes which aspects of such rights might be particularly prone to infringements in the context of big data. The list of fundamental rights at stake is not exhaustive but rather represents the most urgent instances where human rights are challenged as a result of applications of big data currently available. As the available technological solutions used in the context of big data are subject to very fast developments and become more and more ubiquitous, it can be easily imagined that more fundamental rights could be progressively added to the list.

The particular emphasis in the discussion is put on the rights to privacy and data protection. Both rights are in particular implicated at the very first stage of implementing big data technologies – data

[100] See for example: V. Mayer-Schönberger, V. and K. Cukier, *Big Data. A revolution that will transform how we live, work and think*, London: John Murray Publishers, 2013.
[101] Broeders et al., *Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data*, Computer Law & Security Review, Vol. 33 (3), 2017, p. 310.

accumulation, and it has been indicated in the literature that, as such, they might have an enabling effect for ensuring protection of other human rights in the context of new technologies.[102]

Respect for private and family life (Art. 8 ECHR, Art. 7 EU Charter)

Substantive scope of the right

Pursuant to explanations relating to the EU Charter of Fundamental Rights,[103] the right protected under Article 7 of the Charter corresponds to that of the Article 8 of the ECHR. Consequently, in understanding the substantive scope of the right to privacy it is best to turn to the way this right was construed in the case law of the ECtHR.

The umbrella term guaranteeing protection of the right to privacy included in Art. 8.1 ECHR states that everyone has the right to respect for his private and family life, his home and his correspondence. Pursuant to Art. 8.1 ECHR, the scope of the right to privacy is defined very broadly. However, the second paragraph of Art. 8 ECHR sets forth the conditions under which restrictions to privacy can be allowed. The right to privacy in the ECHR is not absolute, but can be restricted if such a restriction is in accordance with the law, necessary for democratic society and aims to pursue one of the enumerated interests: national security, public safety or the economic well-being of the country, prevention of disorder or crime, protection of health or morals, or protection of the rights and freedoms of others.

There is general lack of consensus in legal scholarship as to what the central interests protected by the right to privacy really are and it has been connected to values such as human dignity, liberty and autonomy.[104] The ECtHR does not undertake to resolve this issue and rather takes inherent definitional difficulties as a given in its privacy-related case law: 'The concept of 'private life' is a broad term not susceptible to exhaustive definition, which covers the physical and psychological integrity of a person and can therefore embrace multiple aspects of a person's identity'.[105] The ECtHR notices also the dual nature of the right to privacy, covering not only strictly intimate situations, but also some aspects of social interaction: 'the guarantee afforded by Article 8 of the Convention is primarily intended to ensure the development, without outside interference, of the personality of each individual in his relations with other human beings. There is therefore a zone of interaction of a person with others, even in a public context, which may fall within the scope of 'private life''.[106] In the context of the ECtHR case law it is thus pointless to try to square this notion into one formal category: privacy as a control, privacy as a right to be left alone, etc. It is, however, interesting to look at the spectrum of the interests that the right to privacy aims to protect.

---

[102] For example: M. Oostveen and K. Irion, *The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right?*, Amsterdam Law School Legal Studies Research Paper No. 2016-68, 2016, R. Krotoszynski, *Privacy Revisited: A Global Perspective on the Right to be Left Alone*, Oxford University Press, 2016, D. Solove, *Understanding privacy*, Harvard University Press, 2008.

[103] OJ C 303, 14.12.2007, p. 17–35.

[104] For review of different interests: P. Reagan, *Privacy and the common good: revisited*, in Social Dimensions of Privacy. Interdisciplinary perspectives. Eds. B. Roessler and D. Mokrosinska. Cambridge: Cambridge University Press, 2015, pp. 50-70.

[105] *Axel Springer AG v. Germany* (application no. 39954/08), para. 83.

[106] *Von Hannover v. Germany* (application no. 59320/00), para. 50.

The right to privacy is a multi-faceted right that protects various different interests. The ECtHR has established that the right to privacy not only covers a negative freedom from interference with an individual's life (for example, in case of government surveillance, retention of personal data, illegitimate searches, or publication of private facts), but might also be connected to personal development and establishing relationships with the external world, thus implying certain 'positive' aspects of this right.[107] In this vein the ECtHR recognized for example that certain aspects of reproductive freedom,[108] environmental conditions,[109] awareness of family origins,[110] right to name[111] sexual identity[112] or parental care[113] fell within the scope of the right to privacy.[114] As a result of applying the evolving interpretation by the ECtHR, the scope of the right to privacy has undergone a significant widening. For example, in *Chauvy v. France*[115] and its subsequent case law the ECtHR recognized that the right to reputation should fall within the scope of Art. 8 ECHR. As a result, it expanded the interests protected under this provision to cover not only an individual's internal sphere but also the external one: her public esteem and evaluation in society.[116]

Moreover, although the right to data protection is not as such stipulated in the ECHR, some aspects of the right to data protection have been subsumed within the scope of Article 8 ECHR.[117] The ECtHR considers that wide variety of information as falling within the personal sphere protected by Article 8 ECHR (including, for instance, traffic data from cell phones, emails, voice samples, public photographs, CCTV images). In considering whether certain information falls within the scope of private life, the ECtHR considers different aspects, for example, 'the specific context in which the information at issue has been recorded and retained, the nature of the records, the way in which the records are used and processed and the results that may be obtained".[118] Furthermore, Article 8 ECHR can be infringed even by mere storing of information falling within the scope of Article 8 ECHR[119] and may also concern public information where such information is systematically collected, filed and stored.[120] In its judgment in *Satamedia*[121] the ECtHR underlined that the massive scale of compilation and dissemination of data already available in the public domain can infringe the right to privacy. Moreover, the ECtHR takes into account in its assessment of infringement whether information was used beyond the scope that could have been reasonably

[107] *S and Marper v. The United Kingdom* (applications nos. 30562/04 and 30566/04), para. 66.
[108] For example: *P. and S. v. Poland* (application no. 57375/08), *Evans v. UK* (application no. 6339/05).
[109] For example: Hatton and Others v United Kingdom (application no. 36022/97)
[110] *Backlung v. Finland (application no. 36498/05); Mikulić v. Croatia (application no. 53176/99); Jäggi v. Switzerland (application no. 58757/00).*
[111] For example: *Guillot v. France* (application no. 22500/93).
[112] For example: *Brüggeman and Scheuten v. Germany* (application no. 6959/75), *Schlumpf v. Switzerland* (application no. 29002/06).
[113] For example: B. v. the United Kingdom (application no. 9840/82).
[114] B. van der Sloot, *Privacy as Personality Right: Why the ECtHR's Focus on Ulterior Interests Might Prove Indispensable in the Age of "Big Data"*, 31(80) Utrecht Journal of International and European Law, 2015.
[115] *Chauvy and others v. France* (application no. 64915/01), para. 70.
[116] *Karakó v. Hungary* (application no. 39311/05), paras. 22 and 23.
[117] See: O. Lynskey, *The Foundations of EU Data Protection Law*, OUP, 2015, p.107.
[118] S *and Marper v. The United Kingdom* (applications nos. 30562/04 and 30566/04), para. 67.
[119] *Leander v. Sweden* (application no. 9248/81), para. 48.
[120] *Rotaru v. Romania* (application no. 28341/95), para. 44.
[121] *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland* (application no. 931/13), Grand Chamber.

foreseen by an individual[122] and whether the applicant provided consent for the use of information.[123] In this context, the ECtHR underlines that even if certain use of information was provided for by national laws, such laws must be accessible to the individual and its repercussions clearly foreseeable.[124] In particular, '[i]t follows from well-established case-law that where there has been compilation of data on a particular individual, processing or use of personal data or publication of the material concerned in a manner or degree beyond that normally foreseeable, private life considerations arise'.[125]

In order to trigger the protection granted by Article 8 ECHR, an individual has to attain the status of a victim, meaning that she must indicate to a specific harm suffered which falls within the scope of interests protected by this provision. Thus, pursuant to the ECtHR, the infringement of the right to privacy requires determination of harm. Moreover, such harm cannot be trivial, as pursuant to the *de minimis* principle. Even if certain harm falls under the interests protected by Article 8 ECHR, it also has to attain a certain level of seriousness to be qualified for such protection.[126]

Challenges brought about by the big data technologies

At the outset, it can be noted that big data technologies seem to challenge the right to privacy as set forth in the ECHR and the EU Charter and further developed in the case law of the ECtHR on two different levels. First, big data technologies carry an enormous potential to undermine the very core interests stemming from these rights and connected to the values of individual autonomy and dignity. Second, big data technologies seem to have characteristics that might prevent the effective application of the right to privacy pursuant to the rules established in the ECHR. Moreover, the right to privacy and big data technologies can collide on all specific stages of application of such technologies: the right to privacy can be undermined in the stage of data gathering, processing and potentially also the stage of applying the knowledge derived from processing where it might interfere with the right to privacy.

On the substantive level the first problematic aspect of big data technologies is their volume, in terms of scale of both data gathering (first stage) and further analyses (second stage). Even though data processed in the context of big data technologies do not necessarily contain personal data, aggregate gathering of different data sets might lead to the disclosure of information connected to the most personal spheres. As noted above, pursuant to the ECtHR case law even mere gathering of personal data can fall under the scope of Article 8 ECHR. The ECtHR also noted lawfulness of personal data gathering must be also scrutinized in the light of 'the possible future use of private information retained by the authorities'.[127] In this context, a discussion of the ECtHR on the nature of DNA code as personal data is very illuminating. In *S and Marper v. The United Kingdom* the government argued that DNA code 'is nothing more than a sequence of numbers or a barcode containing information of a purely objective and irrefutable character and that the identification of a subject only occurs in case of a match with another profile in the database'.[128] However, the ECtHR considered that '[w]hile the information contained in the profiles may be considered objective and irrefutable [...], their processing through automated means allows the

[122] *Peck v. The United Kingdom* (application no. 44647/98), para. 62.
[123] *Malone v. The United Kingdom* (application no. 8691/79), para. 84.
[124] *Rotaru v. Romania*, para. 43.
[125] *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland* (application no. 931/13), para. 136.
[126] Article 35 paragraph 3 (b) ECHR, for example : *A. v. Norway*, application no. 28070/06, para. 64.
[127] S *and Marper v. The United Kingdom* (applications nos. 30562/04 and 30566/04), para. 71.
[128] *Ibidem*, para. 74.

Grant Agreement number: 731873

authorities to go well beyond neutral identification'.[129] Thus, the DNA code was to be considered as personal information also on the basis of its potential for revealing future data about an individual. In this sense, in the realm of big data, almost any information, whether personal or not, could be of importance for the purposes of its future impact on the right to private life. Nevertheless, where big data technologies are not based on gathering of data connected in any way to an individual in the initial phase, even if allowing for later aggregation into a specific profile of an individual, such activities would fall outside of the scope of Article 8 ECHR. Similar issues arise, although possibly to a lesser extent, in the context of the right to data protection, which is limited to the scope of personal data. This requires that the data relate to an identified or identifiable natural person. This is not always the case in the context of big data.

Additionally, pursuant to the case law of the ECtHR in the context of the right to reputation, the possibility to maintain a certain public image of oneself is a constitutive aspect of the right to privacy. In this context, data mining and profiling of individuals seem particularly challenging as such technologies might be used to create and maintain certain public profiles of a person and cause stigmatisation. While this issue is connected also to the right of fair trial and the right to non-discrimination (see below), the practice of profiling, augmented by the searchability and permanence of digital records, seems also relevant in the context of an individual's right to relate to others in particular way, to protect certain public identity and esteem inherent in the right of privacy.

The issue of scale is connected to the principle of proportionality, pursuant to which any interference in the right to privacy has to be considered 'necessary in democratic society'. This test 'requires the Court to determine whether the interference complained of corresponded to a pressing social need, whether it was proportionate to the legitimate aim pursued and whether the reasons given by the national authorities to justify it are relevant and sufficient'.[130] This principle clashes with the main assumption behind big data technologies, which is that data is gathered in an indiscriminate manner, without any initial purpose. While big data technologies might prove extremely useful in tackling many pressing social needs, for example, in the area of health care, crime prevention, urban planning, such eventual aims might not be evident at the stage of gathering data and its further processing. In the context of big data, different datasets are used and reused for various different purposes. Thus, the test of proportionality used commonly by the ECtHR to assess whether an infringement of a human right occurred seems not adequate for the big data realm.

Moreover, big data techniques defy the requirements for transparency inherent to the right to privacy as delineated in the case law of the ECtHR. At the stage of data gathering and further processing of data, it was underlined by the ECtHR that 'where there has been compilation of data on a particular individual, processing or use of personal data or publication of the material concerned in a manner or degree beyond that normally foreseeable, private life considerations arise'.[131] Thus, gathering data or further processing of data must be transparent in order to conform to the reasonable expectations of privacy of an individual. Any gathering and further processing of data outside the realm stipulated at the outset might breach such reasonable expectations. Clearly, the practice of gathering as much data as possible for future indeterminate use and reuse applied in the context of big data makes such a requirement of foreseeability futile.

[129] Ibidem, para. 75.
[130] *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland* (application no. 931/13), Grand Chamber, para. 164.
[131] *Ibidem*, para. 136.

Grant Agreement number: 731873

Additionally, where gathering or further processing of data collides with the right to private life, the ECtHR, pursuant to Article 8(2) ECHR will examine if such interference was in accordance with the law. This means that "[t]he law must [...] be formulated with sufficient precision to enable the individual [...] to regulate his conduct.'[132] In the context of secret surveillance or covert intelligence gathering it means specifically that the law has to contain the safeguards against the abuse and arbitrariness, including specifications on: (a) limits on the exercise of powers regarding the storage and use of information; (b) which information could be collected and stored; (c) which categories of people information could be stored about; (d) when surveillance measures were allowed; and (e) limits regarding the length of time for which information could be stored.[133] It is very questionable whether, for example, surveillance and predictive policing technologies based on the application of big data would be able to meet such transparency requirements.

On a procedural level, big data technologies make it difficult for individuals to successfully ascertain their privacy rights in front of courts. First, under current rules recourse to the protection under Article 8 ECHR is available in principle to individuals who can demonstrate their status as a victim,[134] meaning that they have to prove they have personally suffered specific harm against which Article 8 ECHR protects. Moreover, that harm has to be non-trivial. As noted in literature, this focus on an individual might be problematic in the realm of big data.[135] First, in the context of big data, especially during the stages of data gathering and data analyses, it might be extremely difficult to establish what the actual harm is. Second, it might be difficult to prove that an individual suffered such harm. Consequently, '[i]n the big data era, data collection will presumably be so widespread that it is impossible for individuals to assess each data process to determine whether it includes their personal data, if so whether the processing is lawful, and if that is not the case, to go to court or file a complaint'.[136]

Personal data protection (Art. 8 EU Charter)

Substantive scope of the right

The first EU data protection legislation, the Data Protection Directive (DPD) was adopted in 1995. Its double purpose stated in Article 1 is to, on the one hand, ensure protection of fundamental rights and freedoms, in particular the right to privacy with respect to the processing of personal data, and, on the other hand, to facilitate the free flow of data between EU member states. Subsequently, it was included in the EU Charter - as a novelty compared to other international and regional human rights instruments which do not differentiate any separate human right to data protection. In May 2018 the General Data Protection Regulation (GDPR) will enter into force to replace the DPD. As noted by Lynskey, the EU legal framework for data protection can be considered a rights-based regime in the sense that on the one hand it confers specific rights to individuals and on the other hand it has a 'fundamental rights character', meaning that it reflects the underlying concept of the fundamental right to data protection.[137]

---

[132] S *and Marper v. The United Kingdom* (applications nos. 30562/04 and 30566/04), para. 95.
[133] *Rotaru v. Romania*, para. 41, see also: O. Lynskey, *The Foundations of EU Data Protection Law*, OUP, 2015, p. 111.
[134] Article 35(3) ECHR.
[135] B. van der Sloot, *The individual in the Big Data era: Moving towards an agentbased privacy paradigm*, pp. 177-203 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) Exploring the boundaries of Big Data, Amsterdam: Amsterdam University Press, 2016.
[136] B. van der Sloot, S. van Schendel, *International and comparative legal study on Big Data*, The Netherlands Scientific Council for Government Policy Working Paper 20, The Hague 2016.
[137] O. Lynskey, *The Foundations of EU Data Protection Law*, OUP, 2015, p. 38.

Thus, the right to data protection is regulated on the level of primary law (EU Charter) and the level of secondary law (DPD and GDPR respectively). As such, it can be called a human right with a regulatory character[138] since the provisions of the secondary data protection law form a comprehensive regulatory framework, expanding and elaborating the provision in the EU Charter, absent for example in the context of the right to privacy and many other human rights. Specifically, the DPD includes a definition of the term 'personal data', which refers to any information relating to a directly or indirectly identified or identifiable person (Article 2(a) DPD). The scope of this term is thus very broad and in its case law the CJEU found for example that an IP address constitutes personal data pursuant to the provisions of the DPD.[139] Moreover, the CJEU further developed on the concept of identifiability, stating that a natural person is identifiable if there are 'the means which may likely reasonably be used in order to identify the data subject,'[140] which excludes however situations where 'the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant'.[141]

The DPD also specifies the following main principles of data processing, which are slightly expanded under the Article 5(1) of the GDPR:

- lawfulness, fairness and transparency;
- purpose limitation;
- data minimisation;
- accuracy;
- storage limitation;
- integrity and confidentiality;
- accountability.

Moreover the DPD enumerates the grounds under which the processing of personal data is lawful (with the GDPR further developing the conditions for obtaining informed consent from the data subject), including the conditions for the processing of sensitive data (revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health or sex life). Subsequently, the DPD lists the rights available to data subjects in connection to processing their data. The GDPR further expands the list of rights available to individuals in this context and it includes the right to information, right of access, right to rectification, right to erasure, right to restriction of processing, right to data portability, right to object to processing, right to not be evaluated on the basis of automated processing. Additionally, the GDPR dedicates an entire chapter (Chapter IV) to various obligations imposed on the data controllers.

This brief outline of the substantive provisions of the regulatory framework of the fundamental right to data protection created by the DPD (and soon the GDPR) leaves open the question of what is the main interest behind the human right to data protection as a right separate from that to privacy – what is the added value of the right to data protection as such? At the outset it should be noted that the relationship

[138] Gallert, K. de Vries, P. de Hert, and S. Gutwirth, *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, in: B.H.M. Custers, T. Calders, B. Scherme, T. Zarsky (red.) Discrimination and Privacy in the Information Society. nr. 3. Heidelberg: Springer, 2013.
[139] Case C 70/10 *Scarlet Extended*.
[140] Case C 582/14, *Breyer*, para. 48.
[141] *Ibidem*, para. 46.

Grant Agreement number: 731873

between the right to privacy and data protection is blurry and both rights have been often conflated in literature and CJEU case law.[142] Nevertheless, it has been suggested in literature that data protection overlaps considerably with the right to privacy, as they both ensure informational or data privacy, but data protection serves a number of purposes that privacy does not and vice versa.[143] First, it can be noticed that the scope of information considered to be personal data is wider than the information covered by the scope of Article 8 of the EU Charter, where some link to the sphere of private life needs to be established and an individual has to be identified. Second, the right to data protection as regulated in the DPD concerns the wider scope of activities connected to data. Indeed, the scope of the term 'data processing' is extremely wide and covers virtually every imaginable use of personal data at stake. Finally, it has been noted that the right to data protection offers the data subject a wider scope of substantive rights, not all reflected in the case law of the ECtHR concerning the right to privacy (even though the ECtHR progressively includes the catalogue of positive rights, similar to these stipulated in the DPD, attributable to the applicant in the context of the right to privacy).[144]

However, the issue of what is the separate interest covered by the right to data protection remains vague. This point is further exacerbated by the fact that, as has been confirmed by the CJEU in the *Google Spain* case, 'it is not necessary in order to find such a right that [processing of data] causes prejudice to the data subject'.[145] Thus, at first sight, any harm resulting from the infringement of this human right seems to be inherent to the unlawful processing as such. Nevertheless, it has been suggested in literature that the right to data protection has a distinctive normative underpinning, connected to the concept of individual control over the flow of personal data, which aims to prevent different tangible and intangible harms resulting from data processing.[146] In particular, it has been indicated in literature that certain feelings of constant surveillance[147] could be one of the intangible harms connected to forms of data processing in which indiscriminate retention of data takes place.

Finally, it can be noted that the way the EU legislator ensures individual control over personal data through the myriad of specific rights attributed to data subjects 'tilts the balance of interests in data protection law'[148] in their favour. This legislative preference is further exacerbated in the case law of the CJEU, as it seems that the CJEU is willing to give preference to the right to data protection in various cases in which it conflicts with other interests, like in the *Google Spain* case.[149] This is setting the presumption and the precedence in favour of the right to data protection. It might further indicate that the CJEU perceives the right to data protection as vulnerable to hindrances brought about by modern technologies and attempts to mitigate any infringements to this right by strengthening its relative position in relation to other rights.[150]

---

[142] For example: Case C-73/07 *Tietosuojavaltuutettu v Satakunnan Markkinapörssi OY, Satamedia*, Case C-275/06 *Promusicae,* Case C-139/01 *Österreichischer Rundfunk and Others.*
[143] O. Lynskey, *supra*, p. 103.
[144] Ibidem, p. 129.
[145] Case C-131/12 *Google Spain*, para. 96.
[146] Ibidem, p. 179.
[147] Joined Cases C-293/12 and C-594/12, *Digital Rights Ireland*, para. 37.
[148] Ibidem, p. 192.
[149] Case C-131/12 *Google Spain*, para. 81.
[150] M. Jozwiak, *Balancing the Rights to Data Protection and Freedom of Expression and Information by the CJEU: The Vulnerability of Rights in an Online Context*, Maastricht Journal of European and Comparative Law, Special Issue, (3) 2016.

Grant Agreement number: 731873

Challenges brought about by the big data technologies

In the context of data protection, big data and big data technologies seem inherently at odds with what might be considered the main rationale of this right: providing individuals with control over their data. While the right to data protection can be challenged by big data technologies in various different ways, this chapter focuses on the legal aspects of this interaction. Thus, it will be discussed how big data technologies challenge the main legal norms in the DPD and GDPR, aimed at safeguarding the right to data protection. Similarly as in the context of the right to privacy, the right to data protection can potentially clash with big data technologies on all three stages of their application, if only unlawful processing of personal data is involved.

The first challenge concerns the fact that, as discussed above, the focus of the right to data protection is on individual control over personal data. Notably, primary legitimizing ground for data processing is the consent of the data subject. Additionally, the right to data protection grants a set of rights aimed at spurring such control once the data is already being processed, for example, by providing the right to be informed, the right to rectification or the right to access data. The focus on individual rights requires a certain level of transparency regarding by whom and in which manner data is processed. However, in the realm of big data this approach seems flawed as the level of transparency required for an individual to meaningfully exercise the rights granted in the law is practically impossible to attain. As noted by Koops, big data practices 'involve multiple data controllers and processors sharing sets of data, for multiple, not seldom fuzzy, purposes, and increasingly with automated operations on data — think of cloud computing and profiling — that data controllers themselves do not fully understand or know the details of.'[151] Since big data technologies are based in large parts on algorithmic processing and machine learning, data controllers might have very little actual 'control' over different stages of data processing. This makes it even less likely that data subjects would be able to profit from the provisions assuring such control.

Moreover, the principle of purpose limitation warrants that personal data can be collected only for specified legitimate purposes, which in principle constitute the basis for the consent (explicit or otherwise) to process data. Again, the assumption that the data controller determines the purpose of processing *before* the data collection is not practical and realistic from the perspective of big data technologies that often work best when data is not processed with a specific purpose in mind – the analyses are often data-driven and not purpose-driven. Thus, at the moment of data collection it is often not possible for data controllers to envisage what the purpose of the processing might be. In practice, quite often data is used and reused for other purposes.

Similarly, it is striking that while the right to data protection is based on the principle of data minimization, which means that as little data as possible should be collected and personal data must be removed once the goal for which it was gathered has been achieved, the main principle of big data technology is data maximization, as more data makes the algorithmic processing more efficient and allows for discovering more meaningful patters.

Also the principle of data accuracy seems to lose its relevance, as big data technologies are driven by data quantity not quality. Moreover, this principle presupposes that controllers can trace the data back to an individual, while big data technology operates on the level of massive data sets which are not driven by

[151] B.J. Koops, *The trouble with European data protection law*, International Data Privacy Law Advance Access published October 8, 2014.

the individual characteristics but rather certain group features.[152] Given these multiple discrepancies between the current legal framework for data protection and the main characteristics of big data, the question of which regulatory regime might be most adequate in the era of big data arises.

Freedom of expression and information (Art. 10 ECHR, Art. 11 EU Charter)

Substantive scope of the right

Freedom of expression constitutes one of the essential foundations of a democratic society and one of the basic conditions for its progress and for each individual's self-fulfilment.[153] There are three major justifications contemplated in the literature on freedom of expression as the philosophical basis for the right to freedom of speech: democracy, a marketplace of ideas and personal autonomy. Article 11 of the EU Charter and article 10 ECHR both indicate that the right to freedom of expression confers not only the right to speak, but also the right of the audience in general to receive information. Therefore, the notion of freedom of information is concerned with the free flow of speech contained in the public discourse. Such freedom might be thwarted not only when the speaker is silenced, but also when the audience's access to speech is hindered.[154] While the ECtHR has consistently recognized that the public has a right to receive information of general interest, such a right does not grant the general right to access to information. Rather, the right to receive information applies when someone is willing to share it.[155] In its recent case *Cengiz and Others v. Turkey*,[156] the ECtHR found an infringement of right to freedom of expression in a case where a group of individuals was not able to access YouTube as a result of its blanket blocking in Turkey. This finding was made despite the fact that the users who claimed the infringement of their right to freedom of information were not targeted directly by the block of YouTube. The ECtHR found that they had the status of victim in this case, as they were precluded from the access to the only source of information of certain political interest. The right to freedom of information protects the interests of society at large to unfettered access to the sphere of public discourse and the adequate conditions for an individual to freely express opinions. Moreover, the ECtHR supports in its case law the idea of varied and challenging public discourse. The ECtHR often reminds in its case law that the right to freedom of expression "[…] is applicable not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no 'democratic society'"[157].

Additionally, the ECtHR is very aware of the possible chilling effect that certain regulations might have on the right to freedom of speech. The ECtHR underlines that the sanctions imposed for certain speech might

[152] B. van der Sloot, S. van Schendel, *International and comparative legal study on Big Data*, The Netherlands Scientific Council for Government Policy Working Paper 20, The Hague 2016.
[153] For example judgments of the Grand Chamber of the Court in: *Lingens v. Austria* (application no. 9815/82), *Bladet Tromsø and Stensaas v. Norway* (application no. 21980/93), *Axel Springer AG v. Germany* (application no. 39954/08).
[154] Compare for example ECtHR judgment in *Khurshid Mustafa and Tarzibachi v. Sweden* (application no. 23883/06).
[155] For example: *Leander v. Sweden*, *Gaskin v. the United Kingdom* (application no. 10454/83), *Guerra and Others v. Italy* (application no. 14967/89), *Roche v. the United Kingdom* (application no. 32555/96).
[156] *Cengiz and Others v. Turkey* (application nos. 48226/10 and 14027/11).
[157] Handyside v. The United Kingdom (application no. 5493/72), para. 49, repeatedly quoted by the ECtHR in its later cases, for example: Editions Plon v. France, (application no. 58148/00) para. 42, Lindon and others v. France (application no. 21279/02), para. 45.

not only silence the speaker, but also discourage others from expressing their opinions in the future, when the fear of sanctions prevails.[158]

Challenges brought about by the big data technologies

In the context of the right to freedom of expression and access to information, two aspects seem to be particularly challenged by the development of big data technologies. First, the algorithmic processing of information might severely impact people's right to access information. For example, in the context of search engines, the algorithm which determines the sequence of display of pages effectively predetermines which information will be viewed by users. While this feature is very efficient in allowing segregating available information, it might be also the case that a search algorithm gives preference to certain types of content or content providers, thereby diminishing the plurality of available information, especially where increasingly search engines serve as the main interface for acquiring information online.[159]

Moreover, personalisation of information available to different individuals on social media websites, might lead to the creation of so called 'echo chambers' or 'filter bubbles' where users find only information of the ones thinking alike. Such a trend is particularly dangerous having in mind the fact that social media platforms currently become increasingly equivalent to the public sphere.[160] The epitome of the public sphere, as underlined in the case law on the right to freedom of expression, is that one should be exposed to and challenged by various ideas in order to form meaningful opinions.

Another trend connected to algorithmic decision making which challenges the right to freedom of expression appears when certain content is removed from a website due to a purported breach of terms of the website by automated or semi-automated means. For example, platforms like YouTube and Facebook are reported to use automated filtering mechanisms for extremist content.[161] The online platforms face many challenges to ensure that they deliver safe content which does not infringe certain legal norms, concerning for example hate speech, as they are required to verify massive amounts of content – which is virtually impossible using merely human resources. Thus, automation of the process of filtering and deleting online content seems inevitable. However, such automation brings about the risk of over-removal and might infringe the principle of proportionality, as pursuant to Article 10.2 of the ECHR. Any restriction of the right to freedom of expression and access to information must be prescribed by law, pursue one of the legitimate aims foreseen in Article 10.2 ECHR, and must be necessary in a democratic society.

The second aspect of big data technologies challenging the right to freedom of expression is connected to the possible chilling effects caused by the massive processing of data. Big data is based on the phenomenon of 'datafication', meaning that every single piece of information can turned into computable data, and such data is collected on a massive scale, leading to opaque processing which might result in serious repercussions for individuals (for example, decisions on credit scores or insurance rates).

---

[158] For example: Kyprianou v. Cyprus (application no. 73797/01).

[159] B. Wagner, *Study on the Human Rights Dimensions of Algorithms*, Council of Europe, second draft, 20 February 2017.

[160] J. York, *Policing Content in the Quasi-Public Sphere,* Boston, MA: Open Net Initiative Bulletin. Berkman Center, Harvard University, 2010.

[161] See J. Menn, D. Volz, *Exclusive: Google, Facebook quietly move toward automatic blocking of extremist videos,* Reuters, 25 June 2016. http://www.reuters.com/article/us-internet-extremism-video-exclusive-idUSKCN0ZB00M.

Consequently, internet users might feel inhibited not only in their positive freedom in expressing themselves online but also from searching and accessing certain information online. This inhibiting practice might take place in offline environments, as data could be also retrieved from smart devices or CCTV cameras. Thus, oppressive surveillance could also infringe one's right to privacy.

Freedom of assembly and association (Art. 11 ECHR, Art. 12 EU Charter)

Substantive scope of the right

Pursuant to the explanations on the EU Charter, there is an overlap in the meaning of freedom of thought, conscience and religion as set forth in the ECHR and the EU Charter. Moreover, the right to freedom of thought is closely related to the interests protected under the right to freedom of expression, in the way it allows for creating a forum to exchange the ideas and spur open public debate, which is considered a cornerstone of a democratic society.[162]

The ECtHR underlines in particular the importance of access and confrontation with the variety of views and the necessity that governments allow for public expression, 'however shocking and unacceptable certain views or words used may appear to the authorities, and however illegitimate the demands made may be'.[163] It was underlined by the ECtHR that 'pluralism is also built on the genuine recognition of, and respect for, diversity and the dynamics of cultural traditions, ethnic and cultural identities, religious beliefs, artistic, literary and socio-economic ideas and concepts. The harmonious interaction of persons and groups with varied identities is essential for achieving social cohesion.'[164] Thus, the right to freedom of assembly and association aims at guaranteeing individuals a safe platform for interaction.

The right to freedom of assembly and association consists of two dimensions: the right to assembly, for example in the form of manifestation, and the right to association - to create a community regardless of its organizational character. Both these dimensions might impose certain positive obligations on states, obliging them to facilitate the exercising of this right.

Challenges brought about by the big data technologies

It has been underlined that the internet and especially various social media platforms play a vital role in people's freedom to participate in social, political and cultural life and thus can be an important tool enabling organization of assembly and formation of associations.[165] Consequently, individuals should have freedom to use the internet for these purposes.

Similar to the case of freedom of expression, the particular danger stemming from the application of big data technologies might be connected to the use of algorithmic processing and decision-making. For example, as a result of filtering of information available online, certain individuals might be unaware of

[162] *Stankov a.o. v. Bulgaria* (application no. 29221/95), para. 97.
[163] *Ibidem.*
[164] *Gorzelik a.o. v. Poland* (application no. 44158/98), para.92.
[165] Recommendation CM/Rec(2016)5 of the Committee of Ministers to member States on Internet freedom and Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to human rights for internet users.

Grant Agreement number: 731873

the call for assembly or might have restricted access to join an association. Similarly, as a result of preventive policing with the application of algorithms, public authorities might have capacity to prevent the participation of specific groups or individuals in an assembly.[166] Opaque algorithmic filtering processes might lead to divides and fragmentation in society disabling certain 'meeting of minds' necessary for any group interaction which the right to freedom of assembly and association promotes.

Non-discrimination (Art. 14 ECHR, Art. 20 and 21 EU Charter)

Substantive scope of the right

The legal framework of non-discrimination law in both the ECHR and the EU Charter stem from the same general principle of Aristotelian understanding of equality, meaning that similar situations are to be treated in the same manner and different situations are to be treated differently on the basis of their difference. This general principle of equality is specifically set forth in the EU Charter under Article 20. This principle means that there is the obligation not to differentiate between individuals in the same position with respect to relevant grounds (for example, sex) – direct discrimination and that there might be an obligation to differentiate where two individuals are in different situation with connection to that ground (for example, people with disability might require different treatment based on the disability[167]). Setting seemingly neutral rules, measures or criteria for individuals in inherently different situations is thus also treated as discriminatory and is called indirect discrimination. In case of direct discrimination it is necessary to demonstrate certain unfavourable treatment while in case of indirect discrimination it has to be shown that a general neutral rule puts a protected group at particular disadvantage compared with others covered by the rule (for example, a given rule may affect women significantly more than men). Thus, a specific result needs to be shown.[168] In this context, it is very often the case that statistical evidence is taken into account in order to prove discrimination when the measure itself is neutral. The defence to both direct and indirect discrimination exists if there is some objective and reasonable justification for a given treatment, subject to the principle of proportionality. Additionally, some legal instruments specifically mention harassment as a form of discrimination.[169]

Discriminatory behaviour may take different forms and the manner of discrimination is not relevant as such. What is important is whether a treatment puts a certain group protected on relevant ground at a comparative disadvantage. The right to non-discrimination, similar to the right to data protection, can be called a regulatory human right in the EU legal context, as different provisions of secondary legislation specify the rules aimed at facilitating the defence against the discrimination.[170] However, while in the

[166] B. Wagner, *Study on the Human Rights Dimensions of Algorithms*, Council of Europe, second draft, 20 February 2017.
[167] Art. 5 of Directive 2000/78/EC.
[168] 'Put otherwise, the court must be convinced that the only reasonable explanation for the difference in treatment is the protected characteristic of the victim, such as sex or race. The principle applies equally in cases of direct or indirect discrimination', European Union Agency for Fundamental Rights, European Court of Human Rights, Council of Europe, *Handbook on European non-discrimination law,* 2011, p. 213.
[169] Gender Goods and Services Directive, Article 2(d); Gender Equality Directive (Recast), Article 2(1)(d).
[170] R. Gallert, K. de Vries, P. de Hert, and S. Gutwirth, *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, in: B.H.M. Custers, T. Calders, B. Scherme, T. Zarsky (red.) Discrimination and Privacy in the Information Society. nr. 3. Heidelberg: Springer, 2013.

Grant Agreement number: 731873

context of the right to data protection the rights attributed to the data subject are very concrete, under the right to non-discrimination there are no 'fully-fledged subjective rights, but rather guarantees that aim at making action before court successful, thereby ensuring a real judicial efficiency to anti-discrimination principles'.[171] Different measures aimed at facilitating protection against discrimination (such as reversed burden of proof following initial demonstration of probability that discrimination occurred, availability of certain administrative and judicial procedures) aim at assuring that the right to non-discrimination is effective, as disputes over discrimination are hardly ever straightforward and require close scrutiny of the adjudicating body.

There are certain differences in the way the legal systems of the ECHR and the EU Charter regulate the right to non-discrimination. Under the ECHR, Article 14 is treated as an accessory right, albeit of a very broad scope, which might be triggered only where also infringement of some other interests falling within the ambit of the ECHR occurs. In contrast, under EU law the right to non-discrimination is covered by the myriad of provisions of primary law and secondary law, set forth in a number of specific directives stipulating the right to non-discrimination on various grounds and with different scopes of application.

The right to non-discrimination is based on the concepts of specific grounds on which people can be discriminated, for example sexual orientation, age, religious beliefs, disability, race, sex. Both Article 14 ECHR and Article 21 EU Charter contain an open catalogue of such discriminatory grounds. Nevertheless, in the secondary EU legislation the grounds of discrimination are specifically limited to particular fields of application (the Employment Equality Directive[172] prohibits discrimination on the basis of sexual orientation, religious belief, age and disability in the area of employment; the Racial Equality Directive[173] prohibits discrimination on the basis of race or ethnicity in the context of employment, accessing the welfare system and social security, and goods and services; the Gender Goods and Services Directive[174] expands the scope of sex discrimination to also include the area of goods and services apart from employment[175]).

Finally, it is important to note that unlike the ECHR law, under the EU anti-discrimination regime it is not necessary to demonstrate a specific status of a victim in order to challenge a discriminatory measure. Thus, *in abstracto* claims of discrimination are admissible.[176]

[171] *Ibidem*.
[172] Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation (27 November 2000).
[173] Council Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (29 June 2000).
[174] Council Directive 2004/113/EC implementing the principle of equal treatment between men and women in the access to and supply of goods and services (13 December 2004).
[175] Directive 2006/54/EC of the European Parliament and of the Council on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast) (5 July 2006).
[176] Case C-54/07, *Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV*.

Grant Agreement number: 731873

Challenges brought about by the big data technologies

It was noted already in the recitals of the GDPR that data processing may spur the risk of discrimination.[177] Big data 'presents an entirely new stage in the history of discrimination'[178] as it allows connecting massive amounts of data into patterns used as a basis for making decisions about individuals. The right to non-discrimination is most obviously challenged in the context of application of knowledge stemming from profiling either in automated ways or with a human intervention. The algorithmic processing of numerous layers of data is sometimes referred to as a 'black box' – the process is often complex and not transparent and, thus, it may be difficult to detect possible biases which might enter at different stages. First, the way data sets that are fed into the algorithmic processing might be biased leading to discriminatory results. Additionally, algorithms might be designed in a way that uses supposedly neutral proxies, but, when aggregated, lead to discriminatory results. For example, it was shown that based on many different proxies neutral from the point of view of race (such as unemployment history or parental status) the algorithms used in criminal adjudication tended to routinely indicate higher risk of delinquency (information used in probation decisions) for people of colour.[179] Moreover, even if data is not in itself biased and the algorithms are neutral, self-learning machines might intercept the overall bias in society and, for example, connect job offers requiring lesser qualification with the female candidates. Any such differentiation in treatment based on one of the grounds protected under the ECHR and EU legal frameworks, not justified by an objective reason, would be found infringing the human right to non-discrimination.

Discrimination resulting from profiling might be completely unintentional. Sometimes it can be also intentional, covert and hard to detect. For example, parameters for personalization might seem completely neutral but nevertheless reflect bias leading to discrimination. Some parameters might have 'dual valance', meaning that they correlate with both objective and reasonable grounds for differential treatment and prohibited discriminatory ground like race.[180] For such parameters, it might be difficult to establish discrimination pursuant to the norms of the EU non-discrimination legal framework.

It is also important to underline that pursuant to the case law of the CJEU the fact that discriminatory decisions are made on the basis of objective statistical data does not exclude the infringement. Where one of the protected grounds is used in statistical analysis leading to differentiation in treatment, '[w]hat is objectionable (and thus prohibited) in such discrimination is the reliance on characteristics extrapolated from the class to the individual, as opposed to the use of characteristics which genuinely distinguish the individual from others and which may justify a difference in treatment.'[181] Thus, decisions based on analytical tools applying big data would fall into such category of illegal differentiation where, as a result, a person would be treated adversely in connection to one of the protected grounds, despite the seemingly objective character of the analytics.

---

[177] Recitals 75 and 85 GDPR.

[178] P. Hacker, B. Petkova, *Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers*, Northwestern Journal of Technology and Intellectual Property, Forthcoming.

[179] J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias*, ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[180] P. Hacker, B. Petkova, *Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers*, Northwestern Journal of Technology and Intellectual Property, Forthcoming.

[181] Case C-227/04 P, *Lindorfer v. Council*, para. 59, also case C-236/09, *Test-Achats*.

Finally, it should be noted that both the DPD and the GDPR contain provisions that to a certain extent mitigate the risks of algorithmic discrimination as they contain provisions granting the right not to be subject of a decision based 'solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her', subject to certain exceptions.[182] Where such automated decisions concerning an individual are made, an individual should have a right to human intervention in the process and to contest the decision. Such safeguards are also very relevant in the context of the right to fair trial as discussed below. However, it has been argued that from the perspective of big data practices the requirements imposed in Article 22 GDPR are 'unrealistic'[183] as they run counter to the very logic behind big data that implies elimination of the human oversight from the processing. Thus, there seems to be an inherent tension between the strive to capture the benefits of big data analysis and at the same time ensure the fairness of the processes controlled by algorithms and theirs results. Article 22 of the GDPR is perhaps the most salient example of the GDPR's rejection of the big data revolution.[184]

Right to effective remedy and fair trial, presumption of innocence (Art. 6, 13 ECHR, Art. 47, 48 EU Charter)

Substantive scope of the right

There is much overlap between the way the right to effective remedy and fair trial is drafted in the ECHR and the EU Charter. However, several differences render the scope of protection envisaged in the EU Charter slightly wider. Under the ECHR, the presumption of innocence is stipulated within the provision guaranteeing the right to fair trial (Article 6(2) ECHR) and it has the same meaning as the corresponding provision of the EU Charter (Article 48 EU Charter).

With respect to the right of effective remedy, the guarantees conferred by this right are twofold. First, the right to effective remedy (Art. 47 EU Charter) guarantees existence of certain form of redress which can be sought before the court. Under the EU Charter the right to effective remedy protects all the rights granted to an individual within the legal framework of the EU (whereas article 13 ECHR guaranteeing effective remedy has only a subsidiary character relative to the other rights guaranteed under the ECHR).

Second, the right stipulates that the remedy must be effective. Thus it is not enough that there is a mere possibility to access a tribunal – such a possibility must be actually available in practice pursuant to principle of effectiveness. Moreover, while the remedy provided for in legislation can take different forms, it must offer reasonable prospects of success in redressing the infringement.[185] As noted by the ECtHR, the ECHR must be 'interpreted and applied in a manner which renders its rights practical and effective, not theoretical and illusory. Moreover, the Convention is a living instrument which must be interpreted

[182] GDPR, Article 22(1).
[183] A. Rouvroy, '*Of Data and Men': Fundamental Rights and Freedoms in a World of Big Data*, Council of Europe, Directorate General of Human Rights and Rule of Law, p. 11, 2016, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a620
[184] T. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, Seton Hall Law Review, Vol. 47:995, 2017.
[185] *Vučković and Others v. Serbia* (application no. 17153/11), Grand Chamber.

in the light of present-day conditions.'[186] The right to effective remedy ensures that the rights of individuals are effectively protected and promotes the principle of democratic accountability.[187]

The right to fair hearing, a constitutive element of the right to fair trial, guarantees an individual participating in any kind of proceedings before a tribunal (pursuant to provisions of the ECHR the right is limited to civil or criminal proceedings) with certain procedural safeguards promoting the principle of 'equality of arms' which ensures that a person is not disadvantaged in his opportunities to present his case as compared with the adversary. Other safeguards include the right to access to legal representation, public character of the judgment, public hearings, and participation in proceedings.[188]

Finally, one of the core principles of the right to fair trial is the presumption of innocence which means that every person must be presumed innocent and thus cannot be punished, in any way, for the offences suspected of committing until found guilty in a binding judgment (*res judicata*). While the rights to effective remedy, fair hearing and presumption of innocence encapsulate themselves the principle of fairness and are cornerstone of the rule of law they also facilitate individuals in their pursuit to seek redress when other rights are violated.

Challenges brought about by the big data technologies

The use of big data technologies in decision-making processes by private and public actors can inhibit the rights of effective remedy, fair hearing and presumption of innocence in several ways. First, as noted above, the right to effective remedy provides for just this – the remedy which must be effective and not 'illusory'. Drawing on the discussion related to the obstacles in applying current the legal framework for the right to privacy and data protection in the context of big data, the possibility to seek an effective remedy can in many cases be questioned. While the right to data protection set forth in different provisions of the DPD grants an individual a vast spectrum of legal tools to retain control over personal data, these provisions fall short of meeting their aim in practice and a remedy is difficult to envisage. For example, the right to data protection covers only processing of personal data, collecting and mining masses of data. These data are often non-personal data and therefore outside the scope of the DPD despite the fact that such processes can lead to establishing a detailed profile on an individual. Thus, in situations where big data technologies defy the application of the current legal framework for the protection of personal data but nevertheless can lead to infringement of the interests protected by this framework it can be considered that the remedy offered is not effective. The case of *Rotaru v. Romania*[189] is very informative in this context, as the ECtHR considered that the lack of a measure in national law allowing for rectification or erasure of incorrect and secretly gathered information stored by public authorities amounted to a violation of the right to effective remedy in the context of infringement of the right to privacy. *Mutatis mutandis*, the existence of available measures which are incapable of achieving their aims, as in some instances where big data technologies infringe the right to privacy, does not meet the criterion of effectiveness.

[186] *Leyla Sahin v. Turkey* (application no. 44774/98), para. 136.
[187] EU Network of Independent Experts on Fundamental Rights, Commentary of the Charter of Fundamental Rights of the European Union, June 2006.
[188] *Ibidem*.
[189] *Rotaru v. Romania* (application no. 28341/95), Grand Chamber, para. 67.

Grant Agreement number: 731873

Equally, the methods applied in the context of preventive policing can be also considered questionable form the point of view of the rights to fair hearing and the presumption of innocence. In this context, the power of predictive algorithms used by law enforcement agencies could lead to de facto criminalization of certain segments of society and the risks of such technologies are further exacerbated by the fact that the results of big data analytics are often of low quality and not neutral. The first pilot programmes for preventive policing were already launched in Pittsburgh, Pennsylvania, where the algorithms used by the police indicate on maps the places where crimes are likely to happen.[190] Similarly, predictive techniques are used in determining so called 'no-flight' lists aimed at preventing terrorist activities.[191] While such applications can be helpful methods of crime prevention, mitigate certain risks and serving general well-being of society as a whole, from the point of view of human rights they run counter to the principle of the presumption of innocence. The trade-off is problematic in this perspective, because 'big data's promise of increased efficiency, reliability, utility, profit, and pleasure might be seen as the justification for a fundamental jurisprudential shift from our current ex post facto system of penalties and punishments to ex ante preventative measures that are increasingly being adopted across various sectors of society'.[192]

Consumer protection (Art. 38 EU Charter)

Substantive scope of the right

While the legal framework for protection of consumer rights was initially connected within the EU to the sphere of the internal market, it has steadily gained a human rights dimension, where consumers are perceived as inherently more vulnerable agents in the sphere of market transactions in need of protection by the states.[193] In particular, globalisation and technological developments create new realities for customers in which they are exposed to different challenges such as risk of abuse, information asymmetries, and difficulties regarding access to justice.[194] Such new market developments prompted in the EU the idea to confront the challenges facing customers also in the context of the human rights framework and thereby to add the provision in the EU Charter specifying that EU policies shall ensure a high level of consumer protection (Article 38 EU Charter).

The provision on consumer protection is included in the chapter of the EU Charter dedicated to solidarity, which aims to promote human well-being and autonomy.[195] In contrast with previously discussed provisions of the EU Charter, the provision on consumer protection, containing a very broad formulation, could be considered a principle referring to certain policy rather than a subjective right.[196] As such, it does not confer any positive rights to individuals and it cannot be directly claimed in front of any courts, but it

[190] M. Hvistendahl, *Can 'predictive policing' prevent crime before it happen?*, Science Magazine, 28 September 2016, http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens.
[191] S. Panda, *The Procedural Due Process Requirements for No-Fly Lists*, 4 Pierce L. Rev. 121, 2005.
[192] I. Kerr, J. Earle, *Prediction, Preemption, Presumption How Big Data Threatens Big Picture Privacy*, Stanford Law Review Online, September 2015, https://www.stanfordlawreview.org/online/privacy-and-big-data-prediction-preemption-presumption/.
[193] I. Benöhr, *EU Consumer Law and Human Rights*, OUP, 2013.
[194] *Ibidem.*
[195] *Ibidem.*
[196] *Ibidem*.

Grant Agreement number: 731873

might steer member states in the direction of adopting certain legislative instruments.[197] The principle of consumer protection as stipulated in the EU Charter can serve as a guiding tool and cam be applied cumulatively with other EU Charter provisions, reinforcing consumer protection considerations in the legislative instruments or in contractual relationships covered by the scope of the EU Charter.[198]

The sphere of consumer protection is further regulated with specific secondary law tools, not embedded within the framework of human rights, granting a number of rights to consumers.[199]

Challenges brought about by the big data technologies

As noted above, one of the main rationales behind consumer protection is to reinforce the position of consumers vis-à-vis other market players to remedy power imbalances. Big data technologies additionally tilt the power imbalances, giving powerful technological tools to data controllers to influence and control consumer choices, further weakening the position of individuals. From this perspective, the principle of consumer protection, aimed at promoting fairness in commercial relations between consumers and service providers, seems particularly threatened.

Big data technology promises access to many useful services for the customers. However, this is often a at certain price on the side of key values like autonomy or dignity. Power imbalances shaping the contractual relationships between customers and service providers should be taken into account and inform the assessment of their fairness and legality.

3.4 Big data challenges in the context of human rights: the list of issues

The legal framework for subjective human rights coupled with effective judicial enforcement of such rights aims at protecting certain core moral values within the legal realm.[200] However, as discussed above, big data technologies challenge this human rights architecture from many angles. In particular, the legal issues which arise at the nexus between human rights set forth in the EU legal framework and big data technologies could be seen from the perspective of 'within' the legal framework and from 'outside' the legal framework. The former concerns the issues related the application of different human rights in the context of big data and the latter concerns the issues of the more general character, related to application of the legal framework of human rights as a whole.

[197] *Ibidem.*
[198] *Ibidem*.
[199] Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market (Unfair Commercial Practices Directive); Directive 1999/44/EC on certain aspects of the sale of consumer goods and associated guarantees (Sales and Guarantees Directive); Directive 93/13/EEC on unfair terms in consumer contracts (Unfair Contract Terms Directive); Directive 98/6/EC on consumer protection in the indication of the prices of products offered to consumers (Price Indication Directive); Directive 2006/114/EC concerning misleading and comparative advertising (Misleading and Comparative Advertising Directive); Directive 2009/22/EC on injunctions for the protection of consumers' interests (Injunctions Directive).
[200] On this point for example: J. Raz, *Human Rights in the Emerging World Order*, Transnational Legal Theory, pp. 31–47, 2010.

Grant Agreement number: 731873

The list of key legal issues is discussed in broad terms and specified categories are not mutually exclusive. There is necessarily some overlap between different issues as the legal framework for protection of human rights constitutes a coherent whole in which different rights and principles aim at reinforcing one another and at their very heart aim at protecting the same core values. Nevertheless, each issue has a distinctive focus with which challenges of big data are approached that warrants for a separate consideration.

3.4.1 Issues: the application of fundamental rights in the context of big data technologies

Lack of transparency

The issue of transparency revealed to be of particular relevance in context of several human rights discussed. The issue of transparency can be approached from two different angles. First, transparency, i.e. insight into information pertaining to the purposes of data collection, the identity of controller and the kind of personal data processed, not only is intrinsically connected to the right of personal data protection, but also preconditions effective recourse to such rights, including the right to non-discrimination and the right to fair trial and effective remedy. Big data technologies are based on processing of layers upon layers of personal and non-personal data gathered from different sources, allowing for tracing of yet new personal data and new correlations. Combined with the number of different actors implicated in these processes at different stages of applying big data technology, this characterization makes such processes inherently non-transparent. It seems that, without increased transparency into these processes, the essence of the right to data protection and privacy is lost and the effective enjoyment of several other fundamental rights is precluded.

Second, the lack of transparency inherent in big data technologies is at the core of the challenges involving the chilling effects referred to with respect to the right to privacy and the right to freedom of expression, as it creates the sense of constant fear of surveillance and has an inhibitive impact on individual well-being and autonomy.

Vagueness of the concept of harm, lack of individually attributable harm

The issue of lack of transparency is connected with the issue of vagueness of harm resulting from big data applications. Due to the fact that, as explained above, big data technology is not a single linear process, but consists of different stages, with different actors involved and with large amounts of data, the harms connected to these processes can have an incremental character difficult to articulate and difficult to attribute to any given stage or actor. The concept of subjective human rights attributable to each individual presupposes that harm should be specified in order to claim an infringement in front of a court and to safeguard human rights effectively. For example, it might be difficult to pinpoint to specific harm of news personalization, while in aggregate such practices can have impact on the quality and diversity of public discourse and adversely impact the right to access to information. In the same vein, bias of certain algorithmic processing might lead to discriminatory results only in the long run and may not be attributable to one single instance of data processing or algorithmic decision-making. The problem of big data technologies is that many harmless processes might in aggregate lead to infringements of a human right. However, such infringements are difficult to capture pursuant to current framework of human rights.

In connection to the issue of the disperse concept of harm, the issue of lack of individually attributable harm arises. As explained in the previous sections, with certain exceptions in the case law of the ECtHR in the field of data protection[201] and in case of the right to non-discrimination, the ability to make recourse to protection of a human right depends on the individually attributable status of victim. Thus, the harm stemming from a given infringement of a fundamental right must be, pursuant to current human rights framework, suffered specifically by an individual claiming the infringement. In the context of big data, however, it is very often the case that not specific individuals but rather a groups bearing certain common characteristics are targeted.[202] Thus, there arises the issue of conceptualization of a given fundamental right at stake which would allow for protecting harms on a more abstract basis and on a group level.

Proportionality

None of the rights discussed in the previous sections is of absolute character. Thus, in case of conflict with another right or interest it can be limited pursuant to the principle of proportionality. The principle of proportionality, as applied in the case law of the ECtHR,[203] consists of different stages in which the ECtHR verifies whether the limitation of the scope of any given human right was provided for in law, for a legitimate aim and necessary in a democratic society, and in the final stage of the verifying the proportionality *sensu stricto,* the ECtHR balances the conflicting rights. In order to give precedence to one right or another, the ECtHR needs to explicate the values behind these rights, to know which underlying interests would be compromised as a result of an infringement and what is the relative weight of two conflicting values at stake.[204] For example, where the right to data protection of an individual conflicts with the interest of public security the ECtHR needs to establish the relative importance of the values behind these rights and hence the relative gravity of the harms resulting from their limitations.

The application of the principle of proportionality and balancing of conflicting interests at stake might be problematic in the case of big data technologies since the value of their application might have an immediate appeal and the harm might be postponed, vague and dispersed. This calls for coining a list of criteria that may be applicable in cases of such conflicts which would take into account the particularities of big data technologies.

Accountability

The issue of accountability shifts the attention from a rights holder to the duty bearer. The discussion of all the selected human rights revealed that big data technologies challenge to a large extent the possibility of holding any single actor taking part in the stages of big data gathering, processing and decision making accountable, which is in itself problematic from the point of view of the right to effective remedy.

The issue of accountability can be approached by looking at different measures provided for within the current legal framework of human rights which have a potential for holding the infringing actors

[201] For example: *Roman Zakharov v. Russia* (application no. 47143/06).
[202] B. van der Sloot, *The individual in the Big Data era: Moving towards an agent-based privacy paradigm*, pp. 177-203 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) Exploring the boundaries of Big Data, Amsterdam: Amsterdam University Press, 2016.
[203] L. Zucca, *Constitutional Dilemmas: Conflicts of Fundamental Legal Rights in Europe and the USA*, OUP 2009, Chapter 5.
[204] *Ibidem.*

Grant Agreement number: 731873

accountable. For example, the attempts to increase transparency, while greatly cumbersome from a technological point of view in the context of big data, could promote accountability. Additionally, it should be recognized that big data technologies bring about challenges connected with artificial intelligence and self-learning machines and thus ensuring accountability in the context of such automated systems might require entirely novel approaches.[205]

3.4.2 Issues: the fundamental right regulatory framework and big data technologies

Establishing the adequate regulatory framework

While the previous subsection addressed the issues connected to the protection of fundamental rights as currently provided for within different legal instruments in the EU and specific challenges posed by the development of big data technologies, this subsection looks beyond this and tries to suggest possible novel approaches to protecting human rights under pressure.

Given the fact that the current legal framework for protection of human rights displays a number of vulnerabilities in the context of big data applications, finding the proper tools for safeguarding values inherent in these human rights becomes a separate issue. Indeed, in the literature different alternative approaches are suggested for remedying the loopholes in the current system of human rights protection in relation to big data technologies. First, it is suggested that the secondary legal framework for protection of personal data could be modified in different ways. For example, rather than focussing on the stages of data gathering and decision-making, the central stage of data analytics – which is largely unregulated under current framework – could become the focal point for ensuring the protection of human rights at risk by introducing an additional duty of care for those processing personal data.[206] It was also suggested that the values behind the right to data protection would be better protected if the secondary law framework currently in place would be abandoned and instead *sui generis* systems of protection for different kind of data would be adopted.[207] Another method for guaranteeing the protection of values behind human rights in the era of big data suggested in literature calls for shifting the focus from protection of subjective rights for individuals to the obligations of the data controllers, promoting a so-called 'agent-based approach'.[208] As the current regime faces numerous challenges in ensuring the effective protection for the human rights at stake, such alternative approaches could be further explored.

---

[205] J. A. Kroll, H. Huey, S. Barocas, E.W. Felten, J. R. Reidenberg, D. G. Robinson, H. Yu, *Accountable Algorithms*, University of Pennsylvania Law Review, Vol. 165: 633, 2017.

[206] Broeders et al., *Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data*, Computer Law & Security Review, Vol. 33 (3): 309-323, 2017.

[207] J. B. Koops, *The trouble with European data protection law*, International Data Privacy Law Advance Access published October 8, 2014.

[208] B. van der Sloot, *The individual in the Big Data era: Moving towards an agentbased privacy paradigm*, pp. 177-203 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) Exploring the boundaries of Big Data, Amsterdam: Amsterdam University Press, 2016.

The role of private actors in the context of human rights framework

Finally, the issue resulting from the application of the human rights framework to big data technologies which presents many particular challenges from the legal perspective is the role of private actors in this context. The secondary legislation on the right to data protection in both the DPD and the GDPR provides for several provisions which shift the obligation of balancing different fundamental rights and interests on the private actors and it does so without providing any guidance on the relevant criteria to be taken into account, the requirement of substantiating the decision or effective oversight (for example, Article 6(f) GDPR and Article 17(3)(a) GDPR). Opting for such a solution is problematic, as it raises the question of legitimacy of such decisions, especially where obligations of private actors are coupled with the lack of transparency.

Thus, the issue of involvement of private parties in deciding on the issues impacting the human rights of the individuals should be analyzed, taking into account specific incentives that the legal system as a whole might create to tilt the results of such private adjudication.

Table 3, at the beginning of this chapter, provides an overview of all the legal issues identified. As can be seen in the overview, the legal issues are not one-on-one related to the relevant human rights. Rather, the legal issues of big data technologies relate to several or all human rights.

Grant Agreement number: 731873

4. Societal perspective

5. Societal Issue	*Description emphasising societal aspects*
Unequal access	People are not in the same starting position with respect to data and data-related technologies. Certain skills are needed to find one's way in the data era. Privacy policies are usually long and difficult to understand. Moreover, people are usually not able to keep their data out of the hands of parties they don't want to have them.
Normalisation	The services offered to people are selected on the basis of comparisons of their preferences and the preferences of people considered similar to them. People are put into categories whose characteristics are determined by what is most common. There is pressure toward conformity: the breadth of choices is restricted, and pluralism and individuality are pushed back.
Discrimination	People are treated differently based on different individual characteristics or their affiliation to a group. The possibility to reproach people with things they did years ago or to hold people accountable for things they may do in the future affects people's behaviour. The data as well as the algorithms may be incorrect or unreliable, though.
Dependency	People depend on governmental policy for security and privacy purposes. It is considered a misconception that people can be self-governing in a digital universe defined by big data. People choosing not to disclose personal information may be denied critical information, social support, convenience or selection. People also depend on the availability of services provided by companies. It is considered a risk if there are no alternatives to services that are based on the collection or disclosure of personal data.
Intrusiveness	Big data has integrated itself into nearly every part of people's online life and to some extent also in their offline experience. There is a strong sentiment that levels of data surveillance are too intimate but nevertheless many press 'agree' to the countless number of 'terms and conditions' agreements presented to them.
Non-transparency	Algorithms are often like black boxes to people, they are not only opaque but also mostly unregulated and thus perceived as incontestable. People usually cannot be sure who is collecting, processing or sharing which data. Moreover, there are limited means for people to check if a company has taken suitable measures to protect sensitive data.
Abusiveness	Even with privacy regulations in place, large-scale collection and storage of personal data make the respective data stores attractive to many parties including criminals. Simply anonymised data sets can be easily attacked in terms of privacy. The risk of abuse is not limited to unauthorised actors alone but also to an overexpansion of the purposes of data use by authorised actors (e.g. law enforcement, social security).

Table 4 Overview of the societal issues

Grant Agreement number: 731873

Section 4.1 describes a set of key societal issues relevant in the context of big data technologies and their applications, and discusses them in the light of related literature. The set resulted from an in-depth analysis of a larger set of issues extracted from recent research projects. The relative importance of the issues was discussed within the scope of two workshops. The results are summarised in section 4.2. Further issues mentioned in literature but not considered key issues the context of big data technologies and their applications are presented in section 4.3.

4.1 Key issues relevant in the context of big data technologies

Seven issues were selected for further investigation within the scope of the e-SIDES project. The issues are relevant from both a societal and an economic perspective. The term *actors* is used in the following to refer to both individuals and organisations.

As illustrated in **Error! Reference source not found.**2, the key issues are closely connected and each of the issues belongs to one of three groups. The first group of issues, which includes *unequal access*, *normalisation* and *discrimination* focuses on how actors differ or do not differ and how distinctions are made or not made between actors. While the first issue focuses on different starting positions with respect to big data, the second one addresses the neglecting of individual properties and the third one puts the consideration of individual properties in the centre. The second group of issues includes *dependency*, *intrusiveness* and *non-transparency*. The issues of this group focus on the relationship between data subjects and data processors. The first issue puts various forms of dependency in the centre. The second and the third issue focus on the discrepancy between excessive intrusion into the data subjects' affairs on the one side and limited insight into the data processing on the other side. The third group includes only one issue, which is *abusiveness*. Given all the other issues relevant in the context of big data technologies, it is not only likely that data is abused in one form or the other but also that, if abuse happens, its impact may be substantial.

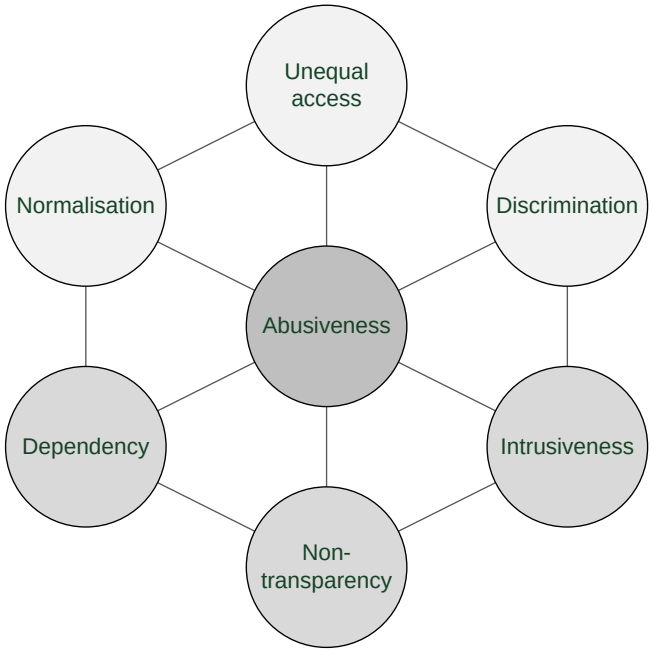

Figure 2 Key societal issues

4.1.1 Unequal access

This issue deals with *unequal access* to data and big data technology, and information asymmetry leading to unequal chances. Not everybody or every organisation is in the same starting position with respect to big data. The digital divide, for instance, refers to inequalities between those individuals who have computers and online access, and those who do not. In big data settings, the digital divide may take larger proportions.[209] Moreover, access to contact data, privacy policies or information about data collection, processing and sharing depends on certain skills. Privacy policies, for instance, are often not only too long but also difficult to understand.[210] The situation, however, is not much different when it comes to the avoidance of access to data. As Hogan and Shepherd state, much of what is involved in keeping personal information out of the hands of intelligence agencies such as the NSA is beyond the ability of most users.[211]

The rather broad concept of the digital divide can easily be transferred to the more specific field of big data. Boyd and Crawford state that *"limited access to big data creates new digital divides"*.[212] Such digital

[209] Custers B.H.M. (2008), The Exclusivity of Ultrafast Communication Networks, *Journal of International Commercial Law and Technology* 3(4): 247-253.
[210] Jonathan A. Obar, "Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management," *Big Data & Society* 2, no. 2 (2015): 8
[211] Hogan and Shepherd, "Information Ownership and Materiality in an Age of Big Data Surveillance," *Journal of Information Policy* 5 (2015): 16
[212] Danah Boyd and Kate Crawford, "Critical questions for big data," *Information, Communication & Society* 15, no. 5 (2012): 673.

Grant Agreement number: 731873

divides are considered to have the potential to create new injustice between groups that have easy access to data and groups that do not.[213]

Relevant inequalities also exist between organisations of different industries, sizes and regional contexts. For instance, inequalities with respect to access to and the replication of data lead to methodological problems and unequal chances for academic and commercial research.[214] An anthropologist working for Facebook and a sociologist working for Google, for instance, have access to data that the rest of the scholarly community does not have.[215] Unsurprisingly, there is also an enormous gap between the developing and the developed world in the utilisation of big data.[216] Data accessibility is much more challenging in developing countries.[217] It is anticipated that future economic and political competitions among countries will be based to a large extent on exploiting the potential of big data.[218]

The need for specific skills together with the widespread lack of these skills among members of the workforce is a key source of unequal access. As the need to manipulate unstructured data increases, the need for more competent data scientists grows.[219] Data scientists, however, who can make sense of big data with a proper understanding of the domain and who are comfortable using analytical tools are not easy to find.[220] The technology required to process big data is either relatively new or became widespread only recently.[221] There is currently a shortage in the job market for skills related to the use, configuration and management of this technology. Even industrialized countries such as those in the European Union (EU) face a shortage of skills with respect to data-related manpower.[222]

Unequal access to data and technology, however, is not only the result of a skills shortage but also linked to the cost and availability of technology, both infrastructure and applications. An organisation wishing to leverage the power of big data may face significant problems related to data complexity and its inherent messiness. Setting up a technology infrastructure for big data analytics requires significant investments in software and hardware.[223] The ability to perform big data analytics is considered as a major differentiator

[213] Ming Xu, Hua Cai and Sai Liang, "Big Data and Industrial Ecology," *Journal of Industrial Ecology* 19, no. 2 (2015)
[214] Ralph Schroeder, "Big Data and the brave new world of social media research," *Big Data & Society* 1, no. 2 (2014): 8
[215] Danah Boyd and Kate Crawford, "Critical questions for big data"
[216] Nir Kshetri, "The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns," *Big Data & Society* 1, no. 2 (2014): 2
[217] Ibid., 14
[218] Xiaolong Jin et al., "Significance and Challenges of Big Data Research," *Big Data Research* 2, no. 2 (2015): 60
[219] In Lee, "Big data: Dimensions, evolution, impacts, and challenges," *Business Horizons* 60, no. 3 (2017): 302
[220] Hans U. Buhl et al., "Big Data," *WIRTSCHAFTSINFORMATIK* 55, no. 2 (2013): 67, Stefan Debortoli, Oliver Müller and Jan Vom Brocke, "Vergleich von Kompetenzanforderungen an Business-Intelligence- und Big-Data-Spezialisten," *WIRTSCHAFTSINFORMATIK* 56, no. 5 (2014): 315, and Veda C. Storey and Il-Yeol Song, "Big data technologies and Management: What conceptual modeling can do," *Data & Knowledge Engineering* 108 (2017): 52
[221] Marco Comuzzi and Anit Patel, "How organisations leverage Big Data: A maturity model," *Industrial Management & Data Systems* 116, no. 8 (2016): 1469
[222] Kshetri, "The emerging role of Big Data in key development issues": 12
[223] Abdulkhaliq Alharthi, Vlad Krotov and Michael Bowman, "Addressing barriers to big data," *Business Horizons* 60, no. 3 (2017): 288

between high-performing and low-performing organisations.[224] It does not only allow organisations to become proactive and forward-looking but also to decrease customer acquisition costs and increase revenues.

Information asymmetry is the obvious result of unequal access to data and technology. Adams and Brückner provide an interesting example that shows that inequalities do not only exist with respect to the utilisation of data but also with respect to the generation of data.[225] They point out that contributing to Wikipedia requires some technical wherewithal that falls outside the skill set of the average Internet user. What is more, beyond the technical skills, negotiating the interaction of editors on Wikipedia requires mastery of a particular jargon and rules of conduct that have evolved in online communities, skills that are similarly not easily acquired by newcomers.

4.1.2 Normalisation

This issue deals with the classification of people and organisations based on categories. *Normalisation* leads to restricted access to information and services. Companies collect and analyse consumers' preference to obtain competitive advantages.[226] Customer relationship strategies, for instance, often prescribe that companies invest more marketing resources in better customers.[227] Consequently, high-value customers receive more marketing than low-value customers. However, many organisations go much beyond that. People are put into categories whose characteristics are determined by what is most common and thus expected to be most likely. The term 'social sorting' is used to refer to the breakdown and categorization of group or person-related raw data into various categories and segments. Adrejevic states that associated with big data, social sorting will range far beyond the marketing realm and allow affecting life chances in increasingly opaque and significant ways.[228]

The literature provides the example of Target, which uses big data analytics through its loyalty card program to track customers' purchasing behaviours and predict their future buying trends.[229] Amazon is another example of a company that is capitalising on big data analytics. Indeed, almost 35% of purchases made on Amazon are generated from personalised purchase recommendations to customers based on

[224] Samuel F. Wamba et al., "Big data analytics and firm performance: Effects of dynamic capabilities," *Journal of Business Research* 70 (2017): 357
[225] Julia Adams and Hannah Brückner, "Wikipedia, sociology, and the promise and pitfalls of Big Data," *Big Data & Society* 2, no. 2 (2015): 1
[226] Pan Liu and Shu-ping Yi, "Investment Decision-Making and Coordination of Supply Chain: A New Research in the Big Data Era," *Discrete Dynamics in Nature and Society* 2016, no. 3 (2016): 2
[227] Charles F. Hofacker, Edward C. Malthouse and Fareena Sultan, "Big Data and consumer behavior: Imminent opportunities," *Journal of Consumer Marketing* 33, no. 2 (2016): 94
[228] Mark Andrejevic, "The Big Data Divide," *International Journal of Communication* 8 (2014)
[229] Wamba et al., "Big data analytics and firm performance": 357

big data analytics.[230] Online recommendation systems categorise customers based on data including the purchase histories of all users.[231]

Amazon uses predictive analytics for targeted marketing to increase customer satisfaction and build company loyalty. The company uses a comprehensive collaborative filtering engine. It analyses what items customers purchased previously, what is in their online shopping cart or on their wish list, which products they reviewed and rated, and what items they searched for.[232] This information is used to recommend additional products that other customers with similar consumption patterns purchased.[233] Recommendation systems depend on data to deliver personalised recommendations.[234] Therefore, smaller companies are disadvantaged in comparison with bigger ones such as Target and Amazon. This links the issues *normalisation* and *unequal access*.

The breadth of choices is restricted and pluralism pushed back. Filter bubbles result when an algorithm selectively guesses what information somebody wants to see based on information about the individual as well as other similar individuals. These are self-reinforcing patterns of narrowing exposure that tend to reduce creativity, learning and connection.[235] Online personalisation is considered to effectively isolate people from a diversity of viewpoints or content. Online recommender systems – built on algorithms that attempt to predict which products or services potential customers will most enjoy consuming – are one family of technologies that potentially suffer from this effect.[236] According to Schroeder, impersonal laws or regularities derived from purchase histories translated in algorithms leave less and less room for individuality.[237]

The attempt of retailing company Target to identify customers in early stages of pregnancy based on their purchasing behaviour is probably one of the best known examples that show how analytics can pose serious privacy risks.[238] The information is valuable as it allows targeted advertising at a critical point in that customer's life when their behaviour is in flux and new habits are formed. In at least one case, a customer's pregnancy was revealed to other members of her family through targeted advertising

[230] Mary J. Wills, "Decisions through data: Analytics in healthcare," *Journal of Healthcare Management* 59, no. 4 (2015)
[231] Xu, Cai and Liang, "Big Data and Industrial Ecology": 206
[232] Matthew J. Mazzei and David Noble, "Big data dreams: A framework for corporate strategy," *Business Horizons* 60, no. 3 (2017): 411
[233] Kshetri, "The emerging role of Big Data in key development issues": 4
[234] Lee, "Big data": 300
[235] Eli Pariser, *The filter bubble: How the new personalized web is changing what we read and how we think* (New York, NY [u.a.]: Penguin Books, 2012)
[236] Tien T. Nguyen et al., "Exploring the filter bubble," in *Proceedings of the 23rd International Conference on World Wide Web*, 677–86 (2014), 677
[237] Schroeder, "Big Data and the brave new world of social media research": 7
[238] Richard Cumbley and Peter Church, "Is "Big Data" creepy?," *Computer Law & Security Review* 29, no. 5 (2013): 603, Sunil Erevelles, Nobuyuki Fukawa and Linda Swayne, "Big Data consumer analytics and the transformation of marketing," *Journal of Business Research* 69, no. 2 (2016): 899, and Abid Mehmood et al., "Protection of Big Data Privacy," *IEEE Access* 4 (2016): 1822

Grant Agreement number: 731873

containing pregnancy-related products. Studies found data consumers are anxious about the collection of personal information via search engines, websites and mobile devices.[239]

Normalisation also happens on an organisational level but is perceived as less critical. In general, large-scale collections of data that allow classifying people and organisations are attractive not only for companies but also for government bodies and criminals.[240] This links the issues *normalisation* and *abusiveness*.

4.1.3 Discrimination

Discrimination is understood as the unfair treatment of people and organisations based on certain characteristics. For a discussion of discrimination issues from an ethical and legal perspective, see the previous chapters. Discrimination may also be a societal issue, as discrimination, particularly when taking place on larger scales may be detrimental to trust among groups in society (social polarisation). Furthermore, discrimination may not always be addressed by legal tools, for instance, in cases of stigmatisation, in which discrimination cannot be enforced. The issue leads to immediate disadvantages and unequal chances. People or groups are treated differently depending on certain characteristics including age, disability, ethnicity or gender.

According to Mayer-Schönberger, the extensive collection of data together with long-term storage, leads to the possibility to reproach people with things they did years ago.[241] This could lead to a situation where people adjust their behavior to be in line with social expectations. Similarly, people may be hold accountable for something they may do in the future. Predictions on the future behavior of people may be taken as justifications for how people are treated today. This may not happen at large scale today but developments in several countries are aiming in this direction.

In most US states, for instance, predictions on the likelihood that somebody will be involved in a violent death over the next 12 months affect the decision if somebody is let out on parole. Machine learning, which is increasingly used for such predictions, however, was found to absorb stereotyped biases towards categories such as race and gender hidden in training data.[242] The number of cities in Europe and beyond, in which the police uses big data analytics to estimate what crimes will be conducted in what neighborhood at what time, is on the rise. Mayer-Schönberger also gives the example of tax authorities that use big data to predict tax evasion. Further examples were discrimination may be the result of big data analytics are no fly lists and credit scores.

---

[239] Matthew S. Eastin et al., "Living in a big data world: Predicting mobile commerce activity through privacy concerns," *Computers in Human Behavior* 58 (2016): 216

[240] Alvaro A. Cárdenas, Pratyusa K. Manadhata and Sreeranga P. Rajan, "Big Data Analytics for Security," *IEEE Security & Privacy* 11, no. 6 (2013): 76

[241] Viktor Mayer-Schönberger, "Big Data - Eine Revolution, die unser Leben verändern wird," [Big data: a revolution that will transform our lives] *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 58, no. 8 (2015): 792

[242] Aylin Caliskan, Joanna J. Bryson and Arvind Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science (New York, N.Y.)* 356, no. 6334 (2017)

The examples available are not limited to criminal prosecution, though. It is also conceivable that somebody does not get access to a specific medical treatment because predictions say that he or she is unlikely to participate in aftercare and rehabilitation actively enough.[243] Insurance companies already calculate insurability and premium level based on individual risk predictions. Personalized pricing examples already show today how discrimination in the context of big data can look like. Harnessing big data collected from customer interactions allows companies to price appropriately and reap the rewards.[244] The US chain of department stores Sears uses big data to help set prices and give loyalty shoppers customized coupons.[245] Orbitz, a travel website, put Apple users at a disadvantage by showing higher priced hotels for customer searches that originated from Apple computers.[246]

Big data technologies to some extent allow concluding initially unknown characteristics from others in the same or other datasets. Recent research indicated that simply anonymised data sets can be easily attacked in terms of privacy.[247] Montjoye et al. collected a 15-month mobility dataset of 1.5 million people.[248] After a simple anonymization operation, a dataset was obtained where the location of an individual was specified hourly with a spatial resolution equal to that given by the carrier's antennas. From the processed data set, they were able to identify a person with 95% accuracy by only four spatial-temporal points.

Discriminating people or groups might make economic sense and is difficult to be detected. Moreover, data or algorithms upon which people are discriminated may be incorrect or unreliable.

4.1.4 Dependency

The *dependency* of people and organisations on organisations and technology leads to a limitation of flexibility. Organisations are strongly dependent on the data as well as the big data technologies they use. Key decisions in fields as critical as health[249] and food[250] more and more rely on big data analytics. Algorithms save time, money and lives but if the they or the data they process are flawed, a vicious circle may be set in motion.[251] A well-known example of the weaknesses of the reliance on informally collected

[243] Ibid.
[244] Walter Baker, Dieter Kiewell and Georg Winkler, "Using big data to make better pricing decisions,", http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions (accessed August 15, 2017)
[245] Lee, "Big data": 300
[246] Eric T. Bradlow et al., "The Role of Big Data and Predictive Analytics in Retailing," *Journal of Retailing* 93, no. 1 (2017): 93
[247] Shui Yu, "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data," *IEEE Access* 4 (2016): 2751
[248] Yves-Alexandre de Montjoye et al., "Unique in the Crowd: The privacy bounds of human mobility," *Scientific reports* 3 (2013)
[249] Prasan K. Sahoo, Suvendu K. Mohapatra and Shih-Lin Wu, "Analyzing Healthcare Big Data With Prediction for Future Health Condition," *IEEE Access* 4 (2016)
[250] Kshetri, "The emerging role of Big Data in key development issues"
[251] John Burn-Murdoch, "The problem with algorithms: magnifying misbehaviour,", https://www.theguardian.com/news/datablog/2013/aug/14/problem-with-algorithms-magnifying-misbehaviour

Grant Agreement number: 731873

data and algorithmic projection is Google Flu Trends, which raised huge scientific optimism about the predictive utility of informally collected data[252] but suffered a serious setback when the predictions for 2013 were shown to be seriously exaggerated[253]. One acknowledged factor for the problems is overconfidence in the veracity of the data as a true sample of reality, rather than a random snapshot in time and the result of algorithmic dynamics.[254]

National statistical offices, for instance, depend on companies to provide them with data.[255] Apart from the fact that direct access to relevant data is hardly possible, the data that is available typically provides only partial information on specific phenomena of interest and does usually not meet any quality standards. The opportunities for large-scale citizen science arise from the ubiquitous networking and computing context and especially the recent growth in the use of mobile devices. The geographic scope of the observational spaces and the varieties of habitats make reliance on trained observers infeasible.[256] Citizen science, however, also poses problems related to dependency.

Individuals depend on governmental policy for security and privacy purposes. Obar, for instance, considers it a misconception that digital citizens can be self-governing in a digital universe defined by big data.[257] Government and corporate regulations for privacy and data protection continue to play a fundamental role in protecting the sensitive aspects of big data.[258] According to Marjani et al., most people are reluctant to rely on systems, which do not provide solid service level agreement conditions regarding theft or misuse of personal information.[259] However, the cost to consumers who choose not to disclose their personal information in today's highly networked world can be substantial . They may be denied critical information, social support, convenience or selection depending on the context. The fact that many online services can only be used after providing requested data (i.e., 'take it or leave it') makes this problem visible.

Technology is essential for big data analytics and must play its part effectively at all stages. Data-intensive organisations such as NHS hospitals in the UK had to stop operating recently after being attacked with ransomware. Big data analytics permanently poses challenges to computation, networking and storage technology.[260] As advancements in technology are necessary to support proper, diverse and timely analytics in times of growing datasets, there is not only a dependence on technology but also on

[252] Jeremy Ginsberg et al., "Detecting influenza epidemics using search engine query data," *Nature* 457, no. 7232 (2009)
[253] David Lazer et al., "Big data. The parable of Google Flu: traps in big data analysis," *Science (New York, N.Y.)* 343, no. 6176 (2014)
[254] Carl Lagoze, "Big Data, data integrity, and the fracturing of the control zone," *Big Data & Society* 1, no. 2 (2014): 5
[255] Enrico di Bella, Lucia Leporatti and Filomena Maggino, "Big Data and Social Indicators: Actual Trends and New Perspectives," *Social Indicators Research* 350, no. 6264 (2016)
[256] Lagoze, "Big Data, data integrity, and the fracturing of the control zone": 7
[257] Obar, "Big Data and The Phantom Public": 2
[258] Ying He et al., "Big Data Analytics in Mobile Cellular Networks," *IEEE Access* 4 (2016): 1993
[259] Mohsen Marjani et al., "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," *IEEE Access* 5 (2017): 5256
[260] Wasim A. Bhat and S. M. K. Quadri, "Big Data promises value: Is hardware technology taken onboard?," *Industrial Management & Data Systems* 115, no. 9 (2015)

technological development. Big data technologies are very complex and need to evolve in order to address issues associated with big data analytics.[261]

According to Symons and Alvarado, big data deals with problems where insights would be practically impossible without the help of computers.[262] Due to the need for unprecedented storage capacity, computing power and efficiency, traditional ways of data modelling lose relevance.[263] Consequently, there is a permanent need for new methods to manage big data for maximum impact and business value.[264]

People and organisations depend on others collecting or processing data, or providing access to data. While cloud computing can be a way to overcome dependency issues related to own technology, it comes with dependency on the cloud provider. If cloud computing is used for big data storage, according to Jain, Gyanchandani and Khare, the data owner loses control over the data.[265]

Switching from one service provider to another is often linked to high costs, if it is possible at all. Lack of interoperability of tools and services as well as lack of data portability are two reasons for dependency in the context of big data. Proprietary and vendor-specific procedures, for instance, were identified as difficulties.[266] Crosas et al. state that *"while there is no lack of big data tools, most of the tools do not communicate or interoperate with each other"*.[267] The lack of interoperability and portability is commonly attributed to the distribution of languages used in big data analytics as well as to the wide distribution of backgrounds and skill sets, disciplines and training. For many types of data or data-related services, there is a limited number of providers and a considerable share of them is based outside the EU.

Business practices as well as security measures can usually not be affected by externals. Without installing proper security mechanisms, confidential information could be transmitted inadvertently to unintended parties, though.[268] For instance, mobile cellular networks have a large amount of sensitive personal information, such as subscriber's names, ID numbers, physical locations, images files, top contacts and passwords.[269] If operators fail to leverage big data in a proper way, big data analytics will bring privacy and security issues to areas such as mobile cellular networks.

---

[261] Jin et al., "Significance and Challenges of Big Data Research": 63 and Storey and Song, "Big data technologies and Management": 56

[262] John Symons and Ramón Alvarado, "Can we trust Big Data? Applying philosophy of science to software," *Big Data & Society* 3, no. 2 (2016): 7

[263] Enrico Barbierato, Marco Gribaudo and Mauro Iacono, "Performance evaluation of NoSQL big-data applications using multi-formalism models," *Future Generation Computer Systems* 37 (2014)

[264] Uthayasankar Sivarajah et al., "Critical analysis of Big Data challenges and analytical methods," *Journal of Business Research* 70 (2017): 274

[265] Priyank Jain, Manasi Gyanchandani and Nilay Khare, "Big data privacy: A technological perspective and review," *Journal of Big Data* 3, no. 1 (2016)

[266] Marjani et al., "Big IoT Data Analytics": 5257

[267] Mercè Crosas et al., "Automating Open Science for Big Data," *The Annals of the American Academy of Political and Social Science* 659, no. 1 (2015)

[268] Lee, "Big data": 301

[269] He et al., "Big Data Analytics in Mobile Cellular Networks": 1986

4.1.5 Intrusiveness

The *intrusion* into peoples' privacy and organisations' business practices leads to a reduction of freedom and autonomy. Big data has integrated itself into nearly every part of people's online life and to some extent also in their offline experience. The behaviour of people including how they live, work and interact is affected by intrusive big data applications. According to Smith, Bennett Moses and Chan, every facet of social life, from healthcare, politics and education to sex, policing and warfare has been touched and modified by the process of digitalisation and datafication.[270] Ever since the Snowden revelations in 2013, there has been a growing awareness of the depth and breadth of the data generated and how it renders citizens into ever more traceable objects of surveillance.[271]

Mayer-Schönberger's examples detailed in the section on *discrimination* also fit here.[272] The possibilities to reproach people with things they did years ago and to hold people accountable for something they may do in the future show that extensive collection of data together with long-term storage can be highly intrusive.

Despite that the information discovered by big data analytics can be very valuable to many applications, people show increasing concern about the other side of the coin, namely the privacy threats posed by data mining.[273] The Target example detailed in the section on *normalisation* also fits here. Target correctly inferred the fact that one of its customers was pregnant by analysing customer data. Due to the high degree of intrusiveness, particularly in the context of biomedical data, there has not only been a debate on the right to know but also the right not to know. [274] The intention behind the right not to know is to protect data subjects from potentially harmful information.

As big data technologies mature, the extensive collection of personal data raises serious concerns for individuals, companies and governments.[275] Without addressing these concerns, individuals may find data analytics worrisome and decide not to contribute personal data that can be analysed later[276]. Companies start to collect and analyse consumers' preference to obtain competitive advantages.[277] However, according to Lee, protecting privacy is often counterproductive to both companies and customers, as big data is a key to enhanced service quality and cost reduction. Therefore, companies and customers need to strike a balance between the use of personal data for services and privacy concerns.

---

[270] Gavin J. D. Smith, Lyria Bennett Moses and Janet Chan, "The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach," *The British Journal of Criminology* 57, no. 2 (2017): 264

[271] Jennifer Pybus, Mark Coté and Tobias Blanke, "Hacking the social life of Big Data," *Big Data & Society* 2, no. 2 (2015) and David Lyon, "Surveillance, Snowden, and Big Data: Capacities, consequences, critique," *Big Data & Society* 1, no. 2 (2014)

[272] Mayer-Schönberger, "Big Data - Eine Revolution, die unser Leben verändern wird": 792

[273] Lei Xu et al., "Information Security in Big Data: Privacy and Data Mining," *IEEE Access* 2 (2014): 1150

[274] Aaro Tupasela and Sandra Liede, "State Responsibility and Accountability in Managing Big Data in Biobank Research: Tensions and Challenges in the Right of Access to Data," in *The Ethics of Biomedical Big Data*, ed. Brent D. Mittelstadt and Luciano Floridi, Law, governance and technology series volume 29 ([Switzerland]: Springer, 2016), 265

[275] Lee, "Big data": 301

[276] Buhl et al., "Big Data": 65 and Lee, "Big data": 301

[277] Liu and Yi, "Investment Decision-Making and Coordination of Supply Chain": 2

At the same time, however, people allow more and more of their actions to be recorded. The 'quantified self' and the 'measured life' are popular names to refer to this phenomenon.[278] More and more people are spending significant portions of their time on social networking.[279] Facebook, for instance, pushes this development with its Facebook Zero initiative that allows mobile users in developing economies to access Facebook free of charge.[280] Research shows that personality traits can be accurately inferred from social network information.[281] At the same time, employers are likely to screen prospective employees through an examination of their social networking profiles. Moreover, as research indicates, social networks such as Facebook but also search engines such as Google and other organisations can effectively manipulate their users.[282]

Immense technology advancement and its intrusion into every aspect of our life are the basic reasons for the creation and growth of big data. If this proliferation and penetration of technology continues, richer and heavier datasets will be created.[283] The proliferation of mobile devices with more and more sensors contributes substantially to this growth.[284]

Despite looking for ways in which they can control their online identity, people often feel caught in what has been referred to as a 'privacy paradox'.[285] There is a strong sentiment that levels of data surveillance are too intimate but nevertheless many press 'agree' to the countless number of 'terms and conditions' agreements, which function largely to legitimise the extraction and monetisation of data.

Since the Snowden revelations, companies such as Google have been using encryption to render the process of data interception more difficult. According to Hogan and Shepherd, the motivation for those companies is not so much to protect privacy in any activist sense but rather to ensure that they do not lose users who rightly fear government monitoring.[286] As mentioned previously, much of what is involved in keeping personal information out of the NSA's hands is beyond the ability of most users. This links the issues *intrusiveness* and *unequal access*. Hogan and Shepherd speak of the Internet as a 'web of surveillance'.

Mobile phones, quantified self devices (wearables, smart glasses, smart bracelets), e-mail, social media networks, community forums and e-commerce sites have increased the opportunity to disclose private information to one or many (whether intentional or not) exponentially.[287] Each of these platforms threatens to expose various levels of an individual's private information. Data is stored for long periods of

[278] Hofacker, Malthouse and Sultan, "Big Data and consumer behavior": 92
[279] Romany F. Mansour, "Understanding how big data leads to social networking vulnerability," *Computers in Human Behavior* 57 (2016): 348 and Jung-Ho Um et al., "Distributed and Parallel Big Textual Data Parsing for Social Sensor Network," *International Journal of Distributed Sensor Networks* 9, no. 12 (2013)
[280] Kshetri, "The emerging role of Big Data in key development issues": 7
[281] Mansour, "Understanding how big data leads to social networking vulnerability": 349
[282] Schroeder, "Big Data and the brave new world of social media research": 3, 5
[283] Bhat and Quadri, "Big Data promises value": 1578
[284] Jochen Binder and Friedemann Weber, "Data Experience — Marktforschung in den Zeiten von Big Data," *Marketing Review St. Gallen* 32, no. 2 (2015): 33
[285] Susan B. Barnes, "A privacy paradox: Social networking in the United States," *First Monday* 11, no. 9 (2006) and Pybus, Coté and Blanke, "Hacking the social life of Big Data": 2
[286] Hogan and Shepherd, "Information Ownership and Materiality in an Age of Big Data Surveillance": 15
[287] Eastin et al., "Living in a big data world": 216

time and the potential to analyse the data or to integrate it with other data grows. General suspicion of public authorities and an insatiable appetite of organisations for ever more data infringe people's freedom.

The impact of the integration of big data and video surveillance is considered to have particular potential for being intrusive. CCTV, body cameras and drones are increasingly used without the consent of the people observed. Supermarkets use video data with face recognition to classify and 'guide' customers.

4.1.6 Non-transparency

The lack of *transparency* of organisational algorithms and business practices is what this issue deals with. It leads to loss of control. Algorithms are often like black boxes to average citizens, they are not only opaque but also mostly unregulated and thus perceived as incontestable. Jain, Gyanchandani and Khare differentiate between active data generation and passive data generation.[288] While active data generation means that the data owner gives the data to a third party, passive data generation refers to the situation where data are produced by a data owner's online actions and the data owner may not know that the data are being gathered by a third party.

People and organisations usually cannot be sure who is collecting, processing or sharing which data. The consumer is often not aware that data collection is taking place at all.[289] This unknowability also reinforces the point being made here, that part of what worries users is that they do not know how much is known about them.[290] It was revealed a number of years ago, already before the explosion of social media, smartphones and tablets, that it would take users an average of 40 minutes a day to read all the privacy policies they encounter.[291] This alone suggests a time management concern associated with self-governance in the big data universe. According to Obar, *"big data analytics is for the most part invisible, managed at distant centers, from behind the scenes, by unnamed powers"*.[292] As a private person, the digital citizen does not know for certain what is going on, or who is doing it, or where they are being carried. Moreover, there are limited means to check if an organisation has taken suitable measures to protect sensitive data.

The NSA but also other agencies around the world intentionally design their surveillance programs to exploit 'backdoors' of private communication platforms like Google and Facebook.[293] Taking advantage of such platforms, intelligence agencies have sought to intercept and collect a living record of human interaction for pervasive surveillance, as manifested in its cross-referenced databases. According to Hogan and Shepherd, *"clicks, uploads, and voices are collected, removed from context, and entrusted to a superhuman algorithm to perpetually aggregate, make sense of, correlate, and render data as evidence"*, for example, to create the no fly lists already mentioned.

---

[288] Jain, Gyanchandani and Khare, "Big data privacy"
[289] Hofacker, Malthouse and Sultan, "Big Data and consumer behavior": 95
[290] Schroeder, "Big Data and the brave new world of social media research": 5
[291] Obar, "Big Data and The Phantom Public": 4
[292] Ibid., 1
[293] Hogan and Shepherd, "Information Ownership and Materiality in an Age of Big Data Surveillance": 7–8

However, there are also many examples of non-transparency beyond what intelligence agencies do. For instance, there are numerous examples of data mining in practice, particularly, in relation to traditional structured datasets such as supermarket loyalty schemes.[294] Again, the perhaps best known example is Target. The example does not only show that companies know a lot about their customers but also that customers are often not aware that this is the case or how far this can already go. This shows links between the issues *non-transparency*, *intrusiveness* and *normalisation*. Over time, even Wikipedia's policies have generated a maze of contradictory rules that are not transparent to newcomers and are enforced by experienced editors often furthering their own interests and agendas.[295]

If the decision maker does not get the data mining results directly from the data miner, he or she would want to know how the results are delivered to him or her and what kind of modification may have been applied to the results, so that he or she can determine whether the results can be trusted.[296] This is why 'provenance' is needed. Of particular concern in this area has been scientific results based on data sources of questionable provenance and integrity such as distributed sensors 'black box social media', where the origin and basis of the data are difficult to determine and the algorithmic bias on the conclusions is difficult to unravel.[297] Lagoze asks in an allusion to citizen science how data or the science that results from those data can be trusted when their provenance is rooted in sources whose own provenance does not conform to 'standard' criteria such as degree, publication record or institutional affiliation?[298]

Some research entails close relations between academics and social media companies who provide access to their data, and to being able to experiment with the platforms. The ethics of academic research may need to be tightened up to provide new guidelines for academic collaboration with commercial platforms, especially, if the line between what part of the research was the company's responsibility and what part was covered by academic ethics review is not clear.[299] There need to be more efforts to specify when access to big data on a new scale enables research that affects many people without their knowledge, and to regulate this type of research – at a minimum making it transparent when such research is being carried out.[300]

Law enforcement is often constrained by a lack of resources of public authorities. Moreover, there is a lack of practical experience with respect to audits including data protection or privacy impact assessments. Concerns with transparency also apply to the growing industry of data brokers. Data brokers acquire detailed and specific information about consumers, analyse it to make inferences about consumers and share the information with clients in a range of industries.[301] All of this activity takes place behind the scenes, often without consumers' knowledge.

[294] Cumbley and Church, "Is "Big Data" creepy?": 603
[295] Adams and Brückner, "Wikipedia, sociology, and the promise and pitfalls of Big Data": 2
[296] Xu et al., "Information Security in Big Data": 1166
[297] Lagoze, "Big Data, data integrity, and the fracturing of the control zone": 5
[298] Ibid., 8
[299] Schroeder, "Big Data and the brave new world of social media research": 3
[300] Schroeder, "Big Data and the brave new world of social media research": 4
[301] Obar, "Big Data and The Phantom Public": 5

Grant Agreement number: 731873

Big data analytics is not only criticised for non-transparency but also praised for its potential to promote transparency and accountability.[302] Transparency involves making information about an entity's operations, structures, and other attributes available to the public. Transparency has gained wide support among state decision-making bodies, international organisations and private companies. Governments can increase transparency by making information available to the public.

4.1.7 Abusiveness

This issue deals with the potential for *abuse* of data and technologies. It leads to control loss and mistrust. In a way, according to Alharthi, Krotov and Bowman, it does not matter how strong or advanced the technical dimension of security is as long as humans are in charge of the data.[303] For instance, many well-known security breaches involved employees simply copying and distributing data to which they had access. Even with privacy regulations in place, large-scale collection and storage of personal information make the respective data stores attractive to many parties including criminals, who, for instance, would like to steal identities.[304]

As already mentioned, recent research indicates that simply anonymised data sets can be easily attacked in terms of privacy by linking two or more datasets.[305] Social networks and the enormous quantities of personal information contained therein do not only constitute a fascinating means of inferring sociological parameters but also a grave risk for security of privacy.[306] The rapid adaptation of malware to social networking sites, for the purposes of social engineering and involuntary surrendering of personal information, shows that the potential for abuse of private information on social networking websites is being exploited.

That digital citizens can be self-governing in a digital universe defined by big data is considered a misconception by Obar.[307] While people face many problems related to the exponential growth of big data, governments seem to champion flawed notice and choice policies.[308] Social media sites can be used to manipulate the audience or customer experiences on an unprecedented scale and with unprecedented accuracy. Politically, according to Schroeder, the main potential abuse of social media could be when authoritarian regimes make use of these techniques in order to browbeat or mollify their citizens.[309] Particularly in countries characterised by conflict, crisis, and weak law enforcement, lack of privacy can quickly become a security risk.[310] Kshetri mentions services such as private investigation, illegal debt collection, asset investigation and kidnapping that can be based on data from the black market.

---

[302] Kshetri, "The emerging role of Big Data in key development issues": 9
[303] Alharthi, Krotov and Bowman, "Addressing barriers to big data": 291
[304] Cárdenas, Manadhata and Rajan, "Big Data Analytics for Security": 76
[305] Xu et al., "Information Security in Big Data": 1155 and Yu, "Big Privacy": 2751
[306] Mansour, "Understanding how big data leads to social networking vulnerability": 348
[307] Obar, "Big Data and The Phantom Public": 2
[308] Ibid.
[309] Schroeder, "Big Data and the brave new world of social media research": 5
[310] Kshetri, "The emerging role of Big Data in key development issues": 15

Grant Agreement number: 731873

Currently, massive amounts of data are collected, processed, analysed and transmitted using high performance parallel and distributed computing systems. However, this big data technology also allows malicious individuals to access high computational power to attack cryptosystems through brute-force attacks.[311] Data as well as big data technologies may be used for illegal purposes or for purposes that fall into a legal grey zone. Data collected to remove security flaws may be used by criminals to take over vulnerable systems. The border between data use and abuse is blurry at times. For instance, it is difficult to check the validity of results of data analyses if they look plausible. Data or algorithms can be manipulated in order to reach desired results. Due to the widespread lack of transparency, such manipulations can hardly be detected. This links the issues *abusiveness* and *non-transparency*.

With respect to non-transparency, it was shown that big data analytics can also increase transparency. Similarly, big data analytics cannot only be abused but it can also help detecting abuse of data and related technologies as well as criminal activities in general. Fraud detection is one of the most visible uses for big data analytics in this regard.[312] Credit card and phone companies have conducted large-scale fraud detection for decades. However, the custom-built infrastructure necessary to mine big data for fraud detection wasn't economical enough to have wide-scale adoption. One of the main impacts from big data technologies is that they're facilitating a wide variety of industries to build affordable infrastructures for security monitoring. According to Cárdenas, Manadhata and Rajan, big data analytics is also particularly suited to become fundamental for advanced persistent threat detection and forensics.

4.2 Relative importance of the issues

Key issues were discussed within the scope of two workshops (see Section 1.2). While the host event of the earlier workshop was a computer ethics conference, the host event of the more recent workshop was a conference focusing on technology management. Consequently, the groups of participants differed a lot in terms of view and understanding.

The more recent workshop focused on the final set of issues described in this document. Towards the end of the workshop, the participants were asked to rate the relevance of the issues in the development of big data technologies on a scale from 1 to 10 (see Appendix A and B). Thereby, a connection between the issues on the one side, and technology-related research and development efforts in the context of big data on the other side was established. Efforts were made to ensure that all participants had understood the issues before the voting started. The results of the voting are shown in Figure 3.

The workshop participants rated *intrusiveness* (average rating 6.0) and *abusiveness* (5.7) as the most relevant issues followed with some distance by *non-transparency* (4.7) and *normalisation* (4.3). The issues *unequal access* (3.9) and *discrimination* (3.9) as well as *dependency* (3.1) were considered less relevant.

[311] Hyun-Ju Jo and Ji W. Yoon, "A New Countermeasure against Brute-Force Attacks That Use High Performance Computers for Big Data Analysis," *International Journal of Distributed Sensor Networks* 11, no. 6 (2015)
[312] Cárdenas, Manadhata and Rajan, "Big Data Analytics for Security": 75

Grant Agreement number: 731873

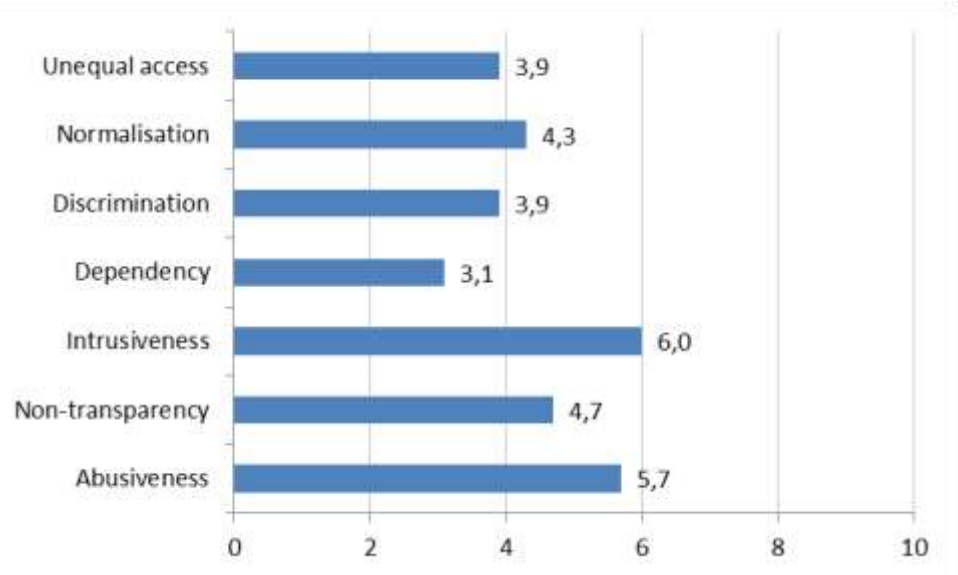

Figure 3 Rating of issues at the most recent e-SIDES workshop

The earlier workshop focused on a slightly larger set of issues. Apart from the issues of the final set, the issues *unfair competition*, *information and power asymmetry* and *labour market transition* were discussed. Within the scope of the workshop, the participants were asked to state to what extent they agree or disagree with a statement saying that the respective issue is relevant and should thus be taken into account when designing big data applications. A scale from 1 (strongly disagree) to 4 (strongly agree) was used. The results are shown in Figure 4.

The workshop participants considered *information and power asymmetry* (average assessment 3.6), *discrimination* (3.4), *intrusiveness* (3.4), *non-transparency* (3.2) and *labour market transition* (3.1) as the most relevant issues. Of the remaining issues, *normalisation* (2.9) and *unfair competition* (2.9) are followed by *dependency* (2.8) and *abusiveness* (2.8) as well as *unequal access* (2.6). With respect to all issues, there is more agreement than disagreement.

The discussion of the results suggested integrating aspects of *information and power asymmetry* into the more general issues *dependency* and *unequal access*. Apart from that, it was decided to merge *unfair competition* with *dependency*. Differing starting positions together with strong but imbalanced interdependencies are characterised by information and power asymmetry and can easily lead to unfair competition. Additionally, the issue *labour market transition* was removed. The labour market transition cannot be directly attributed to big data technologies and their application. It is rather the result of the much broader phenomenon also referred to as digital transformation.

Grant Agreement number: 731873

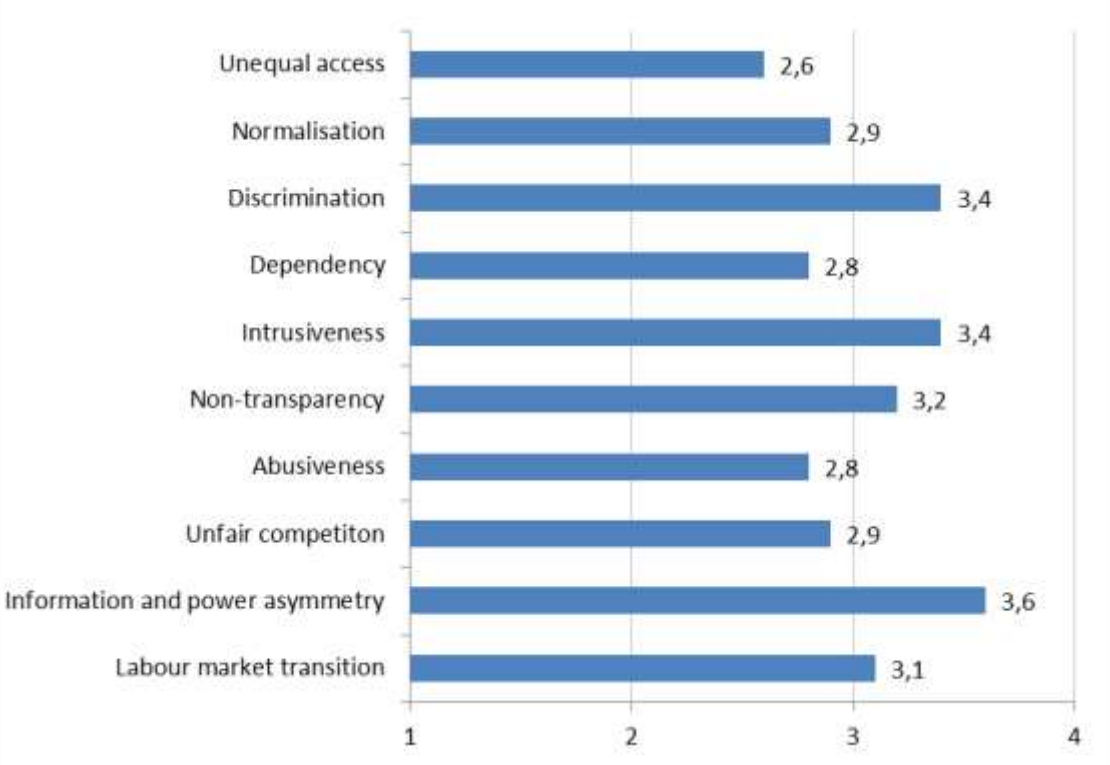

Figure 4 Assessment of issues at the earlier e-SIDES workshop

The comparison of the results of the two workshops shows that views on certain issues differ a lot. To some extent, this may be attributable to the different backgrounds of the participants as well as to changes in the way how the issues are presented and how the votes were carried out. While the results are quite in line with respect to the relatively high relevance of *intrusiveness* and *non-transparency*, *abusiveness* was considered very relevant only by the participants of the more recent workshop. The relative relevance of *discrimination* was only stressed in the earlier workshop. The issues *unequal access* and *dependency* were not considered particularly relevant by the participants of both workshops. This is a bit surprising as aspects of *information and power asymmetry* were integrated into *dependency*.

4.3 Related aspects discussed in the literature

There are several aspects related to the societal and economic issues in the context of big data that did not only receive quite some attention in the literature, but are also considered relevant for the e-SIDES project. Among them are culture, data quality, analytics methodology and visualisation. Understanding them is not only essential to understand the issues, but also to develop means to address them.

Grant Agreement number: 731873

4.3.1 Data culture

The cultural barriers related to big data are considered to be significant and challenging to overcome.[313] This does not only hold for society as a whole, but also for organisations. According to Alharti et al., organisations usually have to adjust their culture to make it supportive of data-driven decision making.[314] This is considered a prerequisite to take full advantage of big data opportunities. Data culture, however, is not only important to seize the opportunities of big data, but also to deal with undesirable implications. People are the key to building and sustaining a data culture. People at all levels must recognize the importance of embracing data and using not only an analytic, but also a socially responsible approach to decision-making. Effective training programmes[315] and actions of the top management of organisations[316] can help to leverage the way users extract and manage data. Nevertheless, it depends a lot on the big data strategy whether data is used in a responsible way, whether disruptive insights are likely, whether a well-founded analytics methodology is used, and whether the cost-benefit ratio and other risks are foreseeable or not.[317] Some big data projects, for instance, have a higher risk of project failure due to unclear problem definitions and the use of emerging technologies.[318] Moreover, the top management of organisations must be aware that performance indicators that often play a key role in the context of data-driven decision-making do not simply measure performance, but may have a lasting effect on it.[319]

4.3.2 Data quality

Apart from data culture, the fitness of data for a specific purpose plays a key role in the context of big data and its applications. Big data applications highly depend on the data that is used.[320] Biases can come in at any step along the data analysis pipeline (see **Error! Reference source not found.**5). Several critical aspects related to data have already been discussed in section 4.1. For instance, it has been mentioned that big data often lacks quality and related standards.[321] Lagoze even defines big data as those data that disrupt fundamental notions of integrity, which is the basis of trust, and force new ways of thinking and doing to re-establish integrity.[322] According to Hofacker et al., data users should not be impressed by the size of a dataset alone and should inquire how the data were sampled and how potential biases may have

[313] Alharthi, Krotov and Bowman, "Addressing barriers to big data": 291
[314] Ibid.
[315] Côrte-Real, Oliveira and Ruivo, "Assessing business value of Big Data Analytics in European firms": 387
[316] Angappa Gunasekaran et al., "Big data and predictive analytics for supply chain and organizational performance," *Journal of Business Research* 70 (2017): 314
[317] Adam Trendowicz, "Analysis of Big Data Potential: How to demonstrate the business value of Big Data?,", IESE-Report 006.17/E (Fraunhofer IESE, 2017), 5
[318] Lee, "Big data": 301
[319] Matthias Becker, *Datenschatten: Auf dem Weg in die Überwachungsgesellschaft?*, 1. Aufl., Telepolis (Hannover: Heise, 2010)
[320] Michael Kläs, Adam Trendowicz and Andreas Jedlitschka, "What Makes Big Data Different from a Data Quality Assessment Perspective? Practical Challenges for Data and Information Quality Research,", IESE-Report 071.15/E (Fraunhofer IESE, 2015), 3–4
[321] Bella, Leporatti and Maggino, "Big Data and Social Indicators" and Lee, "Big data": 301
[322] Lagoze, "Big Data, data integrity, and the fracturing of the control zone": 5

been created by the sampling procedure.[323] Although big data often have complex structures, they still represent only partial observations. Representativeness of data as well as the generalisability of analyses results may be questionable. What Crosas et al. write about research with big data is certainly relevant for big data applications in general. They stress that it should be possible to cite any dataset used for a specific analysis.[324] This is a prerequisite for double-checking results. Repeatability is essential in research but also plays an important role in the context of big data analytics in other contexts. Addressing the undesirable implications of big data analytics makes it necessary that not only the algorithms are reasonably transparent, but also that the fitness of the data analysed is proven. Therefore, it must be possible to find, access and reuse a specific dataset, respecting, of course, the appropriate limitations applied to sensitive data.

Figure 5 Biases can come in at any step along the data analysis pipeline[325]

4.3.3 Analytics methodology

The subjective character of the privacy concept, the lack of a commonly agreed theoretical foundation for privacy in the big data context, the limited scalability and efficiency of privacy-preserving algorithms, and the heterogeneity of data sources make finding or developing the right analytics methodology difficult.[326] Erroneous data and the widespread lack of metadata make big data analytics even more computational intensive. Additionally, they have the potential to make data analytics misleading and difficult to be

[323] Hofacker, Malthouse and Sultan, "Big Data and consumer behavior": 93
[324] Crosas et al., "Automating Open Science for Big Data"
[325] http://www.aolteanu.com/SocialDataLimitsTutorial/index.html
[326] Yu, "Big Privacy": 2760 and Hui Jiang et al., "Energy big data: A survey," *IEEE Access* 4 (2016): 3852

performed in a timely manner and may yield inappropriate results.[327] Data obfuscation is seen as a legitimate means to fight today's pervasive and increasingly intrusive digital surveillance.[328] He et al. state that more advanced algorithms are needed to extract correlations from the data, while allowing different levels of privacy.[329] An alleged cause may be correlated with desirable outcomes, but the correlation could also be due to omitted variables or even reversed causality. If the alleged cause does not affect the outcome, then changing it will not produce the desired change in the outcome variable and the resources spent on the action will be wasted.[330] Large amounts of missing data may cause selection bias and undermine gains in precision afforded by big data, since in multiple regression models, standard statistical software removes observations with missing values.[331] Moreover, security is a key precondition to preserve privacy and to address undesired implications of big data. Apart from secure end-to-end communication, access control is not only particularly relevant but also particularly difficult to maintain in increasingly complex scenarios with many users and highly customised access control levels.[332] Due to the huge number of policies that regulate the access to sensitive data, it can be hard to foresee and predefine all user authorizations, and manually assigning or revoking them when scenario dependent conditions are met.[333]

4.3.4 Visualisation

Due to the large size and high dimension of the data used, visualization is not only an important, but also a difficult task in the context of big data analytics.[334] Visualization solutions need to be compatible with advanced big data analytics frameworks. Additionally, response time is a desirable factor in big data analytics that is also relevant from a visualisation point of view. Presenting results in a manner that is understandable by people without proven skills and experience in data science is highly relevant. A reasonable degree of transparency, a key prerequisite to increase trust in big data, is also relevant with respect to some of the issues identified. Representing key information and knowledge more instinctively and effectively through using different visual formats such as in a pictorial or graphical layout can facilitate striking the balance between the use of personal data and privacy concerns.

[327] Bhat and Quadri, "Big Data promises value": 1590
[328] Finn Brunton and Helen F. Nissenbaum, *Obfuscation: A user's guide for privacy and protest* (Cambridge, Massachusetts: MIT Press, 2015) and D. E. Bakken et al., "Data obfuscation: Anonymity and desensitization of usable data sets," *IEEE Security and Privacy Magazine* 2, no. 6 (2004)
[329] He et al., "Big Data Analytics in Mobile Cellular Networks": 1993
[330] Hofacker, Malthouse and Sultan, "Big Data and consumer behavior": 94
[331] Vera Ehrenstein et al., "Clinical epidemiology in the era of big data: new opportunities, familiar challenges," *Clinical epidemiology* 9 (2017): 248–9
[332] Mehmood et al., "Protection of Big Data Privacy": 1831
[333] Pietro Colombo and Elena Ferrari, "Privacy Aware Access Control for Big Data: A Research Roadmap," *Big Data Research* 2, no. 4 (2015): 149
[334] Marjani et al., "Big IoT Data Analytics": 5258 and Sivarajah et al., "Critical analysis of Big Data challenges and analytical methods": 273

Grant Agreement number: 731873

5 Economic perspective

Economic issue	*Description emphasising economic aspects*
Unequal access	Concerning access to data and technologies, inequalities exist between companies of different industries, sizes and regional contexts. The ability to exploit the potential of big data is a key competitive advantage. The need for specific skills as well as the cost and limited availability of technology are key sources of unequal access.
Normalisation	Companies collect and analyse their customers' preferences, for instance, through loyalty card programs to give personalised purchase recommendations and to obtain competitive advantages. The information gained is valuable as it allows targeted advertising to both private and business customers. The customers' possibilities are restricted, though.
Discrimination	Information about customers and predictions about their behaviour have considerable potential in many sectors from insurance to healthcare and to public administration. Machine learning, which is increasingly used for predictions, however, was found to absorb stereotyped biases towards categories such as race and gender hidden in training data. There is a thin line between legitimate customisation of services and an illegitimate and unfair discrimination of market participants.
Dependency	Companies are strongly dependent on the data as well as the data-related technologies they use. Data is difficult to access, the data available typically provide only partial information and do usually not meet any quality standards. Concerning technology, switching from one service provider to another is often linked to high costs.
Intrusiveness	Protecting privacy is often counterproductive to both companies and customers, as big data is a key to enhanced service quality and cost reduction. Companies and customers need to strike a balance between the use of personal data and privacy concerns. Without addressing these concerns, individuals may find data analytics worrisome.
Non-transparency	Companies know a lot about their customers but customers are often not aware that this is the case. If data from third parties is used, even for companies it is difficult to determine whether the results can be trusted. Moreover, there is a lack of experience with respect to audits including data protection or privacy impact assessments.
Abusiveness	Data as well as data-related technologies may be used for illegal purposes or for purposes that fall into a legal grey zone. Big data analytics, for instance, can be used to manipulate an audience or customer experiences on an unprecedented scale and with unprecedented accuracy. Nevertheless, governments seem to champion flawed notice and choice policies.

Table 5 Overview of the economic issues

Grant Agreement number: 731873

In the e-SIDES project, the economic perspective is relevant for two reasons. First, the issues discussed in Chapter 4 may be relevant for individuals, groups of individuals and society as a whole (societal issues), but they may also affect profit-oriented organisations having business relations with each other as well as with individuals (economic issues). This underlines the close connection of the societal and the economic perspective. Developments of the economy as a whole usually cannot be solely attributed to certain technologies and their degree of privacy-friendliness.

Second, whenever issues are identified and solutions are worked out, the economic viability of the affected business models have to be taken into account. Even if organisations are not inclined or forced to use privacy-preserving big data technologies, they are still very likely to carry out an economic assessment and chose these technologies that promise the highest return on investment in the medium to long run. Consequently, the economic perspective plays a key role when solutions to issues in the context of big data technologies are addressed.

Hence, the societal issues discussed in Chapter 4 are also relevant for the economic perspective. As was explained in Section 1.2, many of the societal issues also include economic aspects and, as such, societal and economic issues cannot always be clearly distinguished. Therefore, the starting point of listing the economic issues are the societal issues. Table 5, at the beginning of this chapter, provides an overview of the issues introduced in section 4.1, now emphasising economic aspects.

Grant Agreement number: 731873

To complement the review of the positions of related projects, the literature analysis and the discussions at the two workshops, relevant studies and reports with an economic focus were investigated in the light of the issues identified. It was found that discussions of economic implications of big data mainly focus on the positive economic potential of the technology and related applications. With respect to economic issues, privacy concerns, the shortage of skilled workforce and technological difficulties received broader attention. The investigation did not result in any important issues that were not yet covered by the issues discussed in section 4.1.

A recent briefing of the European Parliamentary Research Service, for instance, states that big data analytics have the potential to identify efficiencies that can be made in a wide range of sectors and to lead to innovative new products and services, greater competitiveness and economic growth.[335] The briefing cites studies that suggest that companies adopting big data analytics can increase productivity by 5-10% more than companies that do not and that big data practices in Europe could add 1.9% to the GDP between 2014 and 2020. With respect to big data related issues, the briefing highlights the role of protecting privacy, which is the major concern of e-SIDES. Further issues mentioned include data ownership, data localisation, the shortage of skilled workforce and the creation of a new digital divide. Privacy is certainly first and foremost an ethical and a legal issue but it has also been acknowledged as a societal and economic issue. The issues *normalisation*, *discrimination*, *intrusiveness* and *abusiveness* cover aspects of privacy and data protection. The shortage of skilled workforce as well as the creation of a new digital divide are primarily societal and economic issues. The issue *unequal access* clearly covers both aspects, but also *dependency* and *non-transparency* are related to the two issues.

An often cited, but somewhat older report published by the McKinsey Global Institute discusses economic implications of big data at length.[336] For Europe, the authors of the report estimated that government administration could save more than €100 billion through operational efficiency improvements alone by using big data. This estimate does not include big data levers that could reduce fraud, errors and tax gaps. The report states that big data analytics creates value by creating transparency, enabling experimentation, segmenting populations, replacing or supporting human decision making with automated algorithms, and innovating new business models, products and services. The ways how big data creates value mentioned by the McKinsey Global Institute show how closely opportunities and threats are linked in the context of big data. Creating transparency is related to *intrusiveness*, segmenting populations to *normalisation* and *discrimination*, and replacing or supporting human decision making with automated algorithms to *dependency*. With respect to issues that will have to be addressed to capture the full potential of big data, the report mentions data policies, technology (mainly concerned with legacy systems, and incompatible standards and formats), organisational change and talent, access to data, and industry structure. Data policies, as it is framed in the report, deals first and foremost with privacy but takes issues such as security, intellectual property or liability also into account. The connections between the issues identified and privacy have already been discussed in the previous paragraph. Security is covered by *abusiveness*.

[335] Davies, Ron, "Big data and data analytics: The potential for innovation and growth", PE 589.801, European Parliament, European Parliamentary Research Service, Strasbourg, 2016. http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI(2016)589801_EN.pdf

[336] Manyika, James, Michael Chui, Brad Brown et al., "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, 2011. https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Organisational change and talent summarise the lack of a data culture and the shortage of skilled workforce. The skills shortage and how it is covered by the identified issues has already been described. Access to data is covered by *unequal access*. The report stresses that in many cases efficient markets yet have to be set up. Finally, industry structure focuses on aspects such as competitive intensity and performance transparency. The issues *dependency* and *non-transparency* clearly cover what the report summarises as industry structure.

The Science and Technology Committee of the British House of Commons stresses in a recent report that the total amount of global data is predicted to grow 40% per year for the next decade and points out that, properly exploited, this data will be transformative, increasing efficiency, unlocking new avenues in life-saving research and creating new opportunities for innovation.[337] The report, however, acknowledges that there are not only opportunities but also threats. The Science and Technology Committee states that, given the scale and pace of data gathering and sharing, distrust and concerns about privacy and security are often well-founded and must be resolved if the full value of big data analytics is to be realised. As mentioned previously, the issues *normalisation*, *discrimination*, *intrusiveness* and *abusiveness* cover aspects of privacy, data protection and security. The shortage of skilled staff and difficulties related to infrastructure receive particular attention as well. The details of the report also reveal that the Science and Technology Committee of the British House of Commons sees the infrastructure issues very much in line with what was introduced as *unequal access* in section 4.1. The report emphasises problems related to the access to advanced software and hardware, particularly for small companies and researchers, as well as access to data.

In its final report of the European Data Market Study, consultancy firm IDC does not come to fundamentally different results but makes an attempt to put the implications of big data analytics in Europe into figures.[338] The authors of the report describe five economic impacts: increased revenues, reduced costs, enhanced operational efficiency, improved organisational and policy effectiveness, and fostered entrepreneurship with new ventures and cross fertilisation. Apart from that, IDC provides an in-depth analysis of the shortage of skilled workforce as well as citizens' reliance on big data. Both aspects are covered by the issues introduced in section 4.1, particularly *unequal access* and *dependency*. For the EU, IDC calculated 420,000 unfilled data worker positions in 2016. Taken the level of natural unemployment into account, however, IDC does not consider this a very large gap between supply and demand. According to IDC, data workers comprise a wide portfolio of skills and may come from a wide range of disciplines. Due to the dynamics of supply across Europe, it is considered unlikely that there will be major supply problem of data skills in Europe in the future. The gap appears to be very much influenced by demand-supply mismatches by country and industry. With respect to citizens' reliance on big data, the authors of the report state that the health and wellness area is the most relevant one in terms of uptake, impact on behaviour and policy relevance. Health apps and wearables are increasingly used to analyse one's activities and enable citizens to take data-driven decisions on their lifestyle. IDC found out that in

[337] House of Commons, Science and Technology Committee, "The big data dilemma", Fourth Report of Session 2015–16, The Stationery Office, London, 2016. https://publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/468.pdf

[338] IDC, and Open Evidence, "European Data Market", Report for the European Commission, 2017. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=44400

2016 only 4.1% of the EU population used data provided by wearables to drive decisions, varying from 10% in the UK to 0.2% in Romania. While these figures are relatively low, the phenomenon is new and expected to grow strongly.

6 Conclusion

The main aim of this deliverable is to identify and analyse the most relevant ethical, legal, societal and economic issues implicated by the development of big data technologies. In the four previous chapters, each of these four perspectives was taken to identify issues and a list of issues for each perspective was provided. We will not repeat the lists of issues here (we refer to the end of each respective chapter for the tables included), but will draw some general conclusions in this final chapter.

Conclusion 1:

Although there is some **overlap in issues** from the different perspectives, this does not mean that the overlapping issues are the same from each perspective – each perspective simply shows **different aspects of each issue**.

When considering the four lists of issues, it can be observed that there is quite some overlap in these lists. For instance, discrimination is an issue from an ethical, legal, societal *and* economic perspective. This does not mean that the overlapping issues are the same from each perspective – each perspective simply shows different aspects of the issue. This is explained in Table 6, using the example of the issue of discrimination, which appears on all four lists of issues identified in this deliverable. As is explained in Table 6, discrimination from an ethical perspective focuses on the harm to an individual, as opposed to the societal perspective, which focuses on harm to society. At the same time, discrimination as an ethical issue may not (always) be a legal issue. Although some types of discrimination are legally prohibited, such as paying lower wages to women than to men for similar jobs or refusing to hire people because of their religion, some other types of discrimination may be legal, such as refusing to shake hands with women or refusing to befriend coloured people. Discrimination as a societal issue focuses on (harm to) groups of society and society as a whole. Typical examples of discrimination issues on a group level are stigmatisation of particular groups, polarisation in society and social exclusion. Since many of these societal issues merely consist of thoughts and conceptions, they may not always result in actions. Therefore, these societal issues may or may not be illegitimate and may or may not affect individuals. A typical example may be a person who is convinced that all people living in trailers are criminals (stigmatisation), with the exception of his friend, who also lives in a trailer but is 'such a good guy'. This friend may thus not be directly personally affected by these discriminating beliefs and convictions. Discrimination as an economic issue focuses on economic risks, for individuals, companies or society. A typical risk, also from an economic perspective, for organisations is reputational damage caused by exposure of (alleged) discrimination. A typical example of discrimination as an economic issue for individuals is that of price discrimination, for instance, when users of Apple computers pay more for the same products when shopping online than users of regular desktops do.

Grant Agreement number: 731873

Ethical perspective Discrimination as unethical (but not illegal) treatment, causing individual harm Examples: refusing to shake hands with women, refusing to befriend coloured people	**Legal perspective** Discrimination as illegitimate unequal treatment Examples: paying lower wages to women for similar jobs, refusing job applicants because of their religion.
Societal perspective Discrimination causing harm to society Examples: stigmatisation of particular groups in society, polarisation in society, social exclusion	**Economic perspective** Discrimination as an economic risk Examples: reputational damage for organisations, price discrimination (e.g., Apple users paying more when shopping online)

Table 6 The example of discrimination as an issue from each perspective

Conclusion 2:

The list of issues identified is **very extensive**, but **not exhaustive**. The rapid changes in big data technologies call for **periodic updates** of identification of issues.

Although a very broad, comprehensive, multi-method approach was used (see Section 1.2) to map the issues of big data technologies from each of the perspectives that were examined in this deliverable, there is no way of creating an exhaustive overview. There simply does not exist any theoretical framework that is a closed system allowing for an exhaustive approach. Furthermore, because big data technologies and applications are rapidly changing all the time, no approach can be exhaustive. This implies that on the one hand the comprehensive approach taken in this deliverable makes it very likely that the most important issues are actually identified, but on the other hand, this inventory may require periodic updates after some time.

Conclusion 3:

The issues identified are **hard to prioritize**, as this may be **context-dependent** and many issues are **interconnected**.

Although efforts were made to prioritize the issues identified, this is difficult on an abstract level. In a specific context in which big data technologies are used, it may be clearer which issues should be prioritized, but the results of such an exercise may strongly depend on such a context. For instance, from an ethical perspective, depending on the context, different ethical values may have different weights and priorities. Perhaps, in health care setting, non-maleficence of a given treatment may often outweigh privacy. Or, when it comes to driverless cars, autonomy may be less important than human welfare.

Such prioritization is further complicated by the fact that many issues are interconnected. For instance, when profiling takes place and results in price discrimination, some may consider this a privacy issue

Grant Agreement number: 731873

regarding data collection (the data should not be collected in the first place), others may consider it an accountability issue (the data processing was not transparent), and yet others may consider it a justice issue (the decision-making is discriminating and, hence, not just or justifiable).

Conclusion 4:

The issues identified should not only or merely be regarded as problems to be solved, but rather as providing the **goals to strive for**. An attitude of **continuous attention is required** for these issues.

Most or perhaps all of the issues identified are rather complex. Also, they pose challenges that may be hard to solve and it may be hard to determine when they are addressed or solved sufficiently. For instance, reducing the number of cases of discrimination by 50 % (something which would be hard to measure anyway) would be great for everyone in society, but it would still leave many cases unaddressed and, hence, not entirely solve the problem. At the same time, preventing all discrimination from happening may be an unrealistic goal. Hence, a more practical approach would be to strive for addressing the issues *as much as possible*. The issues should, therefore, not be regarded as problems to be solved at once and forever (as in: 'we bought security software, so we no longer need to bother about information security'), but rather as goals to keep striving for (as in: "we bought security software, let's see what the next step of our adversaries is"). An attitude of checking-the-box, whether on legal compliance, data ethics or risk mitigation, is not what is required, as this is perhaps never finished. Rather, an attitude of continuous attention for these issues is called for, including regular updates on the issues themselves.

Bibliography

Adams, Julia, and Hannah Brückner. "Wikipedia, sociology, and the promise and pitfalls of Big Data." *Big Data & Society* 2, no. 2 (2015): 205395171561433.

Alharthi, Abdulkhaliq, Vlad Krotov, and Michael Bowman. "Addressing barriers to big data." *Business Horizons* 60, no. 3 (2017): 285–292.

Baker, Walter, Dieter Kiewell, and Georg Winkler. "Using big data to make better pricing decisions.". http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions (accessed August 15, 2017).

Barbierato, Enrico, Marco Gribaudo, and Mauro Iacono. "Performance evaluation of NoSQL big-data applications using multi-formalism models." *Future Generation Computer Systems* 37 (2014): 345–353.

Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good Frameworks for Engagement* (pp. 44–75). Cambrige University Press.

Baumann, Holgar (2008). "Reconsidering Relational Autonomy. Personal Autonomy for Socially Embedded and Temporally Extended Selves," Analyse and Kritik, 30: 445–468.

boyd, danah and Crawford, Kate, Six Provocations for Big Data (September 21, 2011). A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. Available at SSRN: https://ssrn.com/abstract=1926431 or http://dx.doi.org/10.2139/ssrn.1926431.

Bella, Enrico di, Lucia Leporatti, and Filomena Maggino. "Big Data and Social Indicators: Actual Trends and New Perspectives." *Social Indicators Research* 350, no. 6264 (2016): 1073.

Bhat, Wasim A., and S. M. K. Quadri. "Big Data promises value: Is hardware technology taken onboard?" *Industrial Management & Data Systems* 115, no. 9 (2015): 1577–1595.

Binder, Jochen, and Friedemann Weber. "Data Experience — Marktforschung in den Zeiten von Big Data." *Marketing Review St. Gallen* 32, no. 2 (2015): 30–39.

Boyd, Danah, and Kate Crawford. "Critical questions for big data." *Information, Communication & Society* 15, no. 5 (2012): 662–679.

Bradlow, Eric T., Manish Gangwar, Praveen Kopalle, and Sudhir Voleti. "The Role of Big Data and Predictive Analytics in Retailing." *Journal of Retailing* 93, no. 1 (2017): 79–95.

Braithwaite, John. (2006) "'Accountability and Responsibility through Restorative Justice'". In Public Accountability: Designs, Dilemmas and Experiences, Edited by: Dowdle, M. 33–51. Cambridge: Cambridge University Press.

Beauchamp, T. and Childress, J. (2012) Principles of Biomedical Ethics, 7th edition, New York, Oxford University Press

Booth, A. L. (2008) Environment and Nature: The Natural Environment in Native American Thought in Selin H. (ed.) 'Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures' pp. 809-810, Springer The Netherlands

Brey, P. (2012) Anticipatory Ethics for Emerging Technologies, Nanoethics 6(1), 1-13

Buhl, Hans U., Maximilian Röglinger, Florian Moser, and Julia Heidemann. "Big Data." *WIRTSCHAFTSINFORMATIK* 55, no. 2 (2013): 63–68.

Callicott, B. J. & McRae, J. (Eds.) (2017) Japanese Environmental Philosophy, Oxford University Press

Chuwa L. (2014) Ubuntu Ethics. In: African Indigenous Ethics in Global Bioethics. Advancing Global Bioethics, vol 1. Springer, Dordrecht

Cárdenas, Alvaro A., Pratyusa K. Manadhata, and Sreeranga P. Rajan. "Big Data Analytics for Security." *IEEE Security & Privacy* 11, no. 6 (2013): 74–76.

Comuzzi, Marco, and Anit Patel. "How organisations leverage Big Data: A maturity model." *Industrial Management & Data Systems* 116, no. 8 (2016): 1468–1492.

Côrte-Real, Nadine, Tiago Oliveira, and Pedro Ruivo. "Assessing business value of Big Data Analytics in European firms." *Journal of Business Research* 70 (2017): 379–390.

Crosas, Mercè, Gary King, James Honaker, and Latanya Sweeney. "Automating Open Science for Big Data." *The ANNALS of the American Academy of Political and Social Science* 659, no. 1 (2015): 260–273.

Cumbley, Richard, and Peter Church. "Is "Big Data" creepy?" *Computer Law & Security Review* 29, no. 5 (2013): 601–609.

Custers B.H.M. (2008), The Exclusivity of Ultrafast Communication Networks, *Journal of International Commercial Law and Technology* 3(4): 247-253.

Custers, B.H.M., T. Calders, B. Scherme, T. Zarsky (eds.) Discr*imination and Privacy in the Information Society*. nr. 3. Heidelberg: Springer, 2013.

Custers B.H.M. & Ursic H. (2016), Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection, Int*ernational Data Privacy Law* 6(1): 4-15.

Debortoli, Stefan, Oliver Müller, and Jan Vom Brocke. "Vergleich von Kompetenzanforderungen an Business-Intelligence- und Big-Data-Spezialisten." *WIRTSCHAFTSINFORMATIK* 56, no. 5 (2014): 315–328.

De Hert P., Gutwirth, S. (2006) 'Privacy, data protection and law enforcement. Opacity of the individual and transparency of power' in E. Claes, A. Duff & Gutwirth, S. (eds..), Privacy and the criminal law, Antwerp/Oxford, Intersentia, 61-104

Dendena, Bianca, and Stefano Corsi. "The Environmental and Social Impact Assessment: A further step towards an integrated assessment process." *Journal of Cleaner Production* 108 (2015): 965–977.

Düwell, M. (2017) Human Dignity and the Ethics and Regulation of Technology

Brownsword, R., Scotford, E., Yeung K. (Eds.) The Oxford Handbook of Law, Regulation and Technology, Oxford Handbooks Online

Eastin, Matthew S., Nancy H. Brinson, Alexandra Doorey, and Gary Wilcox. "Living in a big data world: Predicting mobile commerce activity through privacy concerns." *Computers in Human Behavior* 58 (2016): 214–220.

Emery, Mary, and Cornelia Flora. "Spiraling-Up: Mapping Community Transformation with Community Capitals Framework." *Community Development* 37, no. 1 (2006): 19–35.

Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. "Big Data consumer analytics and the transformation of marketing." *Journal of Business Research* 69, no. 2 (2016): 897–904.

Engin, I., & Ruppert, E (2015). *Being Digital Citizens.* London: Rowman & Littlefield International; Hasselbalch, G., & Tranberg, P. (2016). *DATA ETHICS - The New Competitive Advantage*. PubliShare.

Frankel, T. C. & Whoriskey, P. (2016) Tossed aside in the 'white gold' rush Indigenous people are left poor as tech world takes lithium from under their feet, The Washington Post <http://www.washingtonpost.com/graphics/business/batteries/tossed-aside-in-the-lithium-rush/>

Friedman, B. et al. (2006) 'Value Sensitive Design and Information Systems' in (Zhang, N. P. and Galletta, D. eds.) Human-Computer Interaction in Management Information Systems: Foundations, M.E. Sharpe Publishing

Friedman, M. (1998). "Feminism, Autonomy, and Emotion," in Norms and Values: Essays on the Work of Virginia Held, Joram Graf Haber, ed., Lanham, MD: Rowman and Littlefield

Forsberg, E.-M., E. Thorstensen, R. O. Nielsen, and E. de Bakker. "Assessments of emerging science and technologies: Mapping the landscape." *Science and Public Policy* 41, no. 3 (2014): 306–316.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. "Detecting influenza epidemics using search engine query data." *Nature* 457, no. 7232 (2009): 1012–1014.

He, Ying, Fei R. Yu, Nan Zhao, Hongxi Yin, Haipeng Yao, and Robert C. Qiu. "Big Data Analytics in Mobile Cellular Networks." *IEEE Access* 4 (2016): 1985–1996.

Hildebradt, M. de Vries, K. (2013) Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the philosophy of technology, Routledge

Hofacker, Charles F., Edward C. Malthouse, and Fareena Sultan. "Big Data and consumer behavior: Imminent opportunities." *Journal of Consumer Marketing* 33, no. 2 (2016): 89–97.

Hogan, and Shepherd. "Information Ownership and Materiality in an Age of Big Data Surveillance." *Journal of Information Policy* 5 (2015): 6.

Jain, Priyank, Manasi Gyanchandani, and Nilay Khare. "Big data privacy: A technological perspective and review." *Journal of Big Data* 3, no. 1 (2016): 120.

Jin, Xiaolong, Benjamin W. Wah, Xueqi Cheng, and Yuanzhuo Wang. "Significance and Challenges of Big Data Research." *Big Data Research* 2, no. 2 (2015): 59–64.

Jo, Hyun-Ju, and Ji W. Yoon. "A New Countermeasure against Brute-Force Attacks That Use High Performance Computers for Big Data Analysis." *International Journal of Distributed Sensor Networks* 11, no. 6 (2015): 406915.

Jos, P. H. and Tompkins, M. E. (2004) 'The Accountability Paradox in an Age of Reinvention'. Administration and Society, 36(3): 255–81

Kemp, Deanna, and Frank Vanclay. "Human rights and impact assessment: Clarifying the connections in practice." *Impact Assessment and Project Appraisal* 31, no. 2 (2013): 86–96.

Keymolen, E.L.O. (2016). Trust on the line: a philosophycal exploration of trust in the networked era. Erasmus University Rotterdam.

Kshetri, Nir. "The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns." *Big Data & Society* 1, no. 2 (2014): 205395171456422.

Lagoze, Carl. "Big Data, data integrity, and the fracturing of the control zone." *Big Data & Society* 1, no. 2 (2014): 205395171455828.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. "Big data. The parable of Google Flu: traps in big data analysis." *Science (New York, N.Y.)* 343, no. 6176 (2014): 1203–1205.

Lee, In. "Big data: Dimensions, evolution, impacts, and challenges." *Business Horizons* 60, no. 3 (2017): 293–303.

Liu, Pan, and Shu-ping Yi. "Investment Decision-Making and Coordination of Supply Chain: A New Research in the Big Data Era." *Discrete Dynamics in Nature and Society* 2016, no. 3 (2016): 1–10.

Manders-Huits, N. L. J. L., & Van den Hoven, J. (2009). The Need for a Value-Sensitive Design of Communication Infrastructures. In P. Sollie & M. Duwell (Eds.), *Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments.* Boston: Springer.

Mansour, Romany F. "Understanding how big data leads to social networking vulnerability." *Computers in Human Behavior* 57 (2016): 348–351.

Marjani, Mohsen, Fariza Nasaruddin, Abdullah Gani, Ahmad Karim, Ibrahim A. T. Hashem, Aisha Siddiqa, and Ibrar Yaqoob. "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges." *IEEE Access* 5 (2017): 5247–5261.

Mayer-Schönberger, Viktor. "Big Data - Eine Revolution, die unser Leben verändern wird." [Big data: a revolution that will transform our lives] *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 58, no. 8 (2015): 788–793.

Mazzei, Matthew J., and David Noble. "Big data dreams: A framework for corporate strategy." *Business Horizons* 60, no. 3 (2017): 405–414.

Mehmood, Abid, Iynkaran Natgunanathan, Yong Xiang, Guang Hua, and Song Guo. "Protection of Big Data Privacy." *IEEE Access* 4 (2016): 1821–1834.

Mepham, B. (2010) 'The Ethical Matrix as a Tool in Policy Interventions: The Obesity Crisis', in (F-T. Gottwald et al. eds) Food Ethics, Springer Science Business Media, pp. 17-28

McCormick, T. M. (2013). Principles of Bioethics. *Ethics in Medicine*. Retrieved from https://depts.washington.edu/bioethx/tools/princpl.html

Modderkolk, H. (2015). Met big data alleen ga je echt geen aanslagen voorkomen. *Nos.nl*. Retrieved from http://www.volkskrant.nl/buitenland/met-big-data-alleen-ga-je-echt-geen-aanslagen-voorkomen~a4192661/?hash=642ef3fff40bd5faffc383042424afe251927b52;

Moerel L, Prins J.E.J., Hildebrandt, M., Tjong Tjin Tai, T. F. E., Zwenne, G. J. en Schmidt, A. H. J. (2016) Homo Digitalis, Wolters Kluwer Publishing

Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. "Unique in the Crowd: The privacy bounds of human mobility." *Scientific reports* 3 (2013): 1376.

Nguyen, Tien T., Pik-Mai Hui, F. M. Harper, Loren Terveen, and Joseph A. Konstan. "Exploring the filter bubble." In *Proceedings of the 23rd International Conference on World Wide Web*, 677–86. 2014.

Nissenbaum, H. (2010). *Privacy In Context Technology Policy And The Integrity Of Social Life*.

Obar, Jonathan A. "Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management." *Big Data & Society* 2, no. 2 (2015): 205395171560887.

Omer, T. and Polonetsky, J. "Privacy in the Age of Big Data: A Time for Big Decisions." February 2, 2012. 64 Stan. L. Rev. Online 63. http://www.stanfordlawreview.org/online/privacy-paradox/big-data (last visited June 28, 2012)

Owen, Richard, John Bessant, and Maggy Heintz, eds. *Responsible Innovation*. Chichester, UK: John Wiley & Sons, Ltd, 2013.

Pariser, Eli. *The filter bubble: How the new personalized web is changing what we read and how we think*. New York, NY [u.a.]: Penguin Books, 2012.

Proceedings of the 23rd International Conference on World Wide Web. 2014.

Page, K. (2012). The four principles - Can they be measured and do they predict ethical decision-making? *BMC Medical Ethics*, *13*(10).

Price, J., Price, D., Williams, G., & Hoffenberg, R. (1998). Changes in medical student attitudes as they progress through a medical course. *J Med Ethics*, *24*(2), 110.

Pybus, Jennifer, Mark Coté, and Tobias Blanke. “Hacking the social life of Big Data.” *Big Data & Society* 2, no. 2 (2015): 205395171561664.

Rawls, J. (1971). A Theory of Justice, Revised edition (1999) Cambridge, MA: Harvard University Press. Christman, John and Joel Anderson, eds. (2005). Autonomy and the Challenges to Liberalism: New Essays, New York: Cambridge University Press.

Richards, N. M., & King, J. H. (May, 19, 2014). Big Data Ethics. *Wake Forest Law Review*, 2014

Sahoo, Prasan K., Suvendu K. Mohapatra, and Shih-Lin Wu. “Analyzing Healthcare Big Data With Prediction for Future Health Condition.” *IEEE Access* 4 (2016): 9786–9799.

Schroeder, Ralph. “Big Data and the brave new world of social media research.” *Big Data & Society* 1, no. 2 (2014): 205395171456319.

Sivarajah, Uthayasankar, Muhammad M. Kamal, Zahir Irani, and Vishanth Weerakkody. “Critical analysis of Big Data challenges and analytical methods.” *Journal of Business Research* 70 (2017): 263–286.

Smith, Gavin J. D., Lyria Bennett Moses, and Janet Chan. “The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach.” *The British Journal of Criminology* 57, no. 2 (2017): 259–274.

Storey, Veda C., and Il-Yeol Song. “Big data technologies and Management: What conceptual modeling can do.” *Data & Knowledge Engineering* 108 (2017): 50–67.

Strandburg, K. (2014). Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good : Frameworks for Engagement*. Cambridge University Press.;

Symons, John, and Ramón Alvarado. “Can we trust Big Data? Applying philosophy of science to software.” *Big Data & Society* 3, no. 2 (2016): 205395171666474.

Synder, J. E., & Gauthier, C. C. (2008). The Underlying Principles of Ethical Patient Care. In *Evidence-based Medical Ethics* (pp. 11–17). Humana Press

Tischner, J. (2005) The Ethics of Solidarity, Retrieved on 22nd of August from < http://www.tischner.org.pl/Content/Images/tischner_3_ethics.pdf>

Um, Jung-Ho, Chang-Hoo Jeong, Sung-Pil Choi, Seungwoo Lee, Hwan-Min Kim, and Hanmin Jung. “Distributed and Parallel Big Textual Data Parsing for Social Sensor Network.” *International Journal of Distributed Sensor Networks* 9, no. 12 (2013): 525687.

Vallor, S. (2017) Technology and the Virtues: A philosophical guide for a future worth wanting, New York, Oxford University Press, pp. 120-121

Vanclay, Frank. “International Principles For Social Impact Assessment.” *Impact Assessment and Project Appraisal* 21, no. 1 (2003): 5–12.

Vanclay, Frank, Ana M. Esteves, Ilse Aucamp, and Daniel M. Franks. “Social impact assessment: Guidance for assessing and managing the social impacts of projects.”. https://www.iaia.org/uploads/pdf/SIA_ Guidance_Document_IAIA.pdf (accessed August 11, 2017).

Van den Hoven, J. (2007). ICT and Value Sensitive Design. In V. Goujon, P.; Lavelle, S.; Duquenoy, P.; Kimppa, K.; Laurent (Ed.), *IFIP International Federation for Information Processing, The Information Society: Innovations, Legitimacy, Ethics and Democracy* (Vol. 233, pp. 67–72). Boston: Springer.

Wamba, Samuel F., Angappa Gunasekaran, Shahriar Akter, Steven J.-f. Ren, Rameshwar Dubey, and Stephen J. Childe. “Big data analytics and firm performance: Effects of dynamic capabilities.” *Journal of Business Research* 70 (2017): 356–365.

Waltho, S. (2006). Response to Westin and Nilstun. *Health Care Analysis*, *14*(2).

Westin, L., & Nilstun, T. (2006). Principles help to analyse but often give no solution - secondary prevention after a cardiac event. *Health Care Analysis*, *14(2).*

Wills, Mary J. "Decisions through data: Analytics in healthcare." *Journal of Healthcare Management* 59, no. 4 (2015): 254–262.

Xu, Lei, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren. "Information Security in Big Data: Privacy and Data Mining." *IEEE Access* 2 (2014): 1149–1176.

Xu, Ming, Hua Cai, and Sai Liang. "Big Data and Industrial Ecology." *Journal of Industrial Ecology* 19, no. 2 (2015): 205–210.

Yu, Shui. "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data." *IEEE Access* 4 (2016): 2751–2763.

Zook, M., Barocas, S., Crawford, K., Keller, E., Goodman, A., Hollander, R., … Pasquale, F. (2017). Ten simple rules for responsible big data research. *Computational Biology*, *13*(3), 1–10.

Case law of the European Court of Human Rights

A. v. Norway (application no. 28070/06)
Axel Springer AG v. Germany (application no. 39954/08)
B. v. the United Kingdom (application no. 9840/82)
Backlung v. Finland (application no. 36498/05)
Bladet Tromsø and Stensaas v. Norway (application no. 21980/93)
Brüggeman and Scheuten v. Germany (application no. 6959/75)
Cengiz and Others v. Turkey (applications nos. 48226/10 and 14027/11)
Chauvy and others v. France (application no. 64915/01)
Editions Plon v. France, (application no. 58148/00)
Evans v. UK (application no. 6339/05)
Gaskin v. the United Kingdom (application no. 10454/83)
Gorzelik a.o. v. Poland (application no. 44158/98)
Guerra and Others v. Italy (application no. 14967/89)
Guillot v. France (application no. 22500/93)
Handyside v. The United Kingdom (application no. 5493/72)
Hatton and Others v United Kingdom (application no. 36022/97)
Jäggi v. Switzerland (application no. 58757/00)
Karakó v. Hungary (application no. 39311/05)
Khurshid Mustafa and Tarzibachi v. Sweden (application no. 23883/06)
Kyprianou v. Cyprus (application no. 73797/01)
Leander v. Sweden (application no. 9248/81)
Leyla Sahin v. Turkey (application no. 44774/98)
Lindon and others v. France (application no. 21279/02)
Lingens v. Austria (application no. 9815/82)
Malone v. The United Kingdom (application no. 8691/79)
Mikulić v. Croatia (application no. 53176/99)
P. and S. v. Poland (application no. 57375/08)
Peck v. The United Kingdom (application no. 44647/98)
Roche v. the United Kingdom (application no. 32555/96)
Roman Zakharov v. Russia (application no. 47143/06)
Rotaru v. Romania (application no. 28341/95)
S and Marper v. The United Kingdom (applications nos. 30562/04 and 30566/04)
Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland (application no. 931/13)
Schlumpf v. Switzerland (application no. 29002/06)
Soering v. the United Kingdom (application no. 14038/88)
Stankov a.o. v. Bulgaria (application no. 29221/95)
Tyrer v. United Kingdom (application no. 5856/72)
Von Hannover v. Germany (application no. 59320/00)
Vučković and Others v. Serbia (application no. 17153/11)

Case law of the Court of Justice of the EU

Case 29/69 *Stauder*, 12 November 1969
Case C-139/01 *Österreichischer Rundfunk and Others*, 20 May 2003
Case C-144/04 *Mangold v. Helm*, 22 November 2005
Case C-227/04 P *Lindorfer v. Council*, 11 September 2007
Case C-275/06 *Promusicae*, 29 January 2008
Case C-54/07 *Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV*, 10 July 2008
Case C-73/07 *Tietosuojavaltuutettu v Satakunnan Markkinapörssi OY, Satamedia*, 16 December 2008
Case C-555/07 *Kücükdeveci v. Sweden*, 19 January 2010
Case C-236/09 *Test-Achats*, 1 March 2011
Case C 70/10 *Scarlet Extended*, 24 November 2011
Case C-617/10 *Åkerberg Fransson*, 26 February 2013
Case C-399/11 *Melloni*, 26 February 2013
Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland*, 8 April 2014
Case C-131/12 *Google Spain*, 13 May 2014
Case C 582/14 *Breyer*, 19 October 2016

Grant Agreement number: 731873

Appendix A Workshop questions of ethical and legal issues

First Workshop (CEPE/Ethicomp Conference 2017 in Turin)

To what extent do you agree/disagree with the following statements?
Informed consent is a relevant issue and should be taken into account when designing big data applications
Purpose limitation is a relevant issue and should be taken into account when designing big data applications
Sensitive data is a relevant issue and should be taken into account when designing big data applications
Harm of processing is a relevant issue and should be taken into account when designing big data applications
Solidarity is a relevant issue and should be taken into account when designing big data applications
Trust is a relevant issue and should be taken into account when designing big data applications
Autonomy is a relevant issue and should be taken into account when designing big data applications
Bias is a relevant issue and should be taken into account when designing big data applications
Opacity is a relevant issue and should be taken into account when designing big data applications
Moral responsibility is a relevant issue and should be taken into account when designing big data applications

Grant Agreement number: 731873

Second workshop (ICE/IEEE Conference 2017 in Madeira)

Which of the following issues should be prioritized in the development of big data technologies in order to minimise negative impacts?
Lack of fully informed consent for use of personal data
Ineffective purpose limitation of the exploitation of personal data
Blurring of the concept of sensitive data
Harm done through processing not clearly understood
Solidarity - undermining social cohesion by using personal data profiling
Loss of trust because of dependency on Big data providers
Risk of reducing autonomy through the manipulation of individual choices
Profiling, categorization and correlation create Bias in society
Avoidance of moral responsibility when decision making is machine generated.

Grant Agreement number: 731873

Appendix B Workshop questions of societal and economic issues

First workshop (CEPE/Ethicomp Conference 2017 in Turin)

To what extent do you agree/disagree with the following statements?
Unequal access is a relevant issue and should be taken into account when designing big data applications
Normalization is a relevant issue and should be taken into account when designing big data applications
Discrimination is a relevant issue and should be taken into account when designing big data applications
Dependency is a relevant issue and should be taken into account when designing big data applications
Intrusiveness is a relevant issue and should be taken into account when designing big data applications
Non-transparency is a relevant issue and should be taken into account when designing big data applications
Abusiveness is a relevant issue and should be taken into account when designing big data applications
Unfair competition is a relevant issue and should be taken into account when designing big data applications
Information and power asymmetry is a relevant issue and should be taken into account when designing big data applications
Labor market transition is a relevant issue and should be taken into account when designing big data applications

Grant Agreement number: 731873

Second workshop (ICE/IEEE Conference 2017 in Madeira)

Which of the following issues should be prioritized in the development of big data technologies in order to minimise negative impacts?
The diversity of people and organizations in terms of capabilities, resources, and access to data and technologies leads to unequal chance
The reduction of people and organizations to a norm leading to limitation of choice
Unfair treatment of people and organizations based on certain characteristics leads to unequal chances
The dependency of people and organizations from organizations and technology leading to limitation of flexibility
The intrusion into the peoples' privacy and organizations' business practices leading to reduction of freedom
The lack of transparency of organizational algorithms and business practices leading to control loss
The potential for abuse of data and technologies leading to control loss and deep mistrust

Grant Agreement number: 731873